

Hubris, benefits and minefields of human cloning

One man's absurdities are a gift to witless stereotypers of scientists. They are also an unwelcome stimulus to much needed consideration of the realities and implications of an uncomfortable technology.

Eleven years ago, a group of Argentine scientists wrote to *Nature* complaining about the way in which a US research group had chosen to carry out unauthorized experiments in their country using a rabies-vaccine recombinant virus on cattle (see *Nature* 324, 610; 1986). The event caused a worldwide furore, not least because of the suspicion that the US researchers were taking advantage of a more relaxed regulatory system to carry out experiments that would have been refused permission at home. The experiments were quickly terminated. But they remain deeply entrenched in institutional memories, and have played an important part in ensuring that close — perhaps excessive — attention is given to safety issues in debates about the UN Biodiversity Convention.

Argentina's experience is directly relevant to the issues facing politicians and civil servants as they struggle to respond to the challenge thrown down last week by Richard Seed, the Chicago fertility enthusiast who claims to be about to carry out experiments in human cloning. One of the many arrogant statements produced by Seed was that, if prevented from carrying out these experiments in the United States, he would merely do so abroad, possibly in Mexico. The naivety of his belief that he will be able to find any responsible government prepared to offer him shelter from the opprobrium of the world's political and religious leaders (or indeed the doggedness of the world press) is matched only by the hubris with which Seed appears to have brushed aside the daunting technical obstacles that confront any attempt at safe and successful human cloning.

Whatever Seed's real capacities may be (and it is perhaps significant that at least one influential figure in the American Physical Society has been at pains to distance himself from any description of Seed as a physicist), with friends like him, human cloning has little need of enemies. Indeed, even if the most determined opponents of cloning had sought to invent a caricature of the 'mad scientist' encapsulating their greatest fears, they would probably have had qualms about writing some of the things that Seed has said for fear of destroying all credibility. The possible benefit of Seed behaving in the way that he has is that it has reawakened an awareness of the need for political action that had lost its way since an initial flurry last year, in the wake of the birth of the cloned lamb Dolly (see page 218). The undoubted disadvantage is that his excesses have enhanced polarization in the debate, encouraging a reflex response that demands sweeping bans rather than the determined reflection that the issue deserves.

Distinctions

There are important distinctions to be drawn within the cloning debate that must be addressed if the many beneficial possibilities opened up by work at the Roslin Institute and elsewhere are not to be stifled (as some of the rules on, for example, embryo research and the use of fetal tissue in the United States have already done). These include the regeneration of diseased or damaged tissue and body parts made possible by more thorough knowledge of the techniques needed to manipulate the expression of the genes responsible for

organ development. To include all such possibilities under the simplistic and now dangerously emotive label 'human cloning' undermines any reasoned debate on where a line should be drawn between what is and what is not acceptable, and in what circumstances.

Such distinctions are essential. Regrettably, to some there is as much arrogance behind the factual statement that a human clone is little different from an identical twin as there is in the demand that no human cloning research should be permitted. Again, deliberate and considered reflection is required. Think, for a moment, of the implications of the knowledge that an individual is a genetically identical copy of a parent. One does not need to make hand-waving appeals to abstract ideas of 'human dignity', or even delve into the realms of psychoanalysis, to appreciate that although some are relaxed about such a possibility, that feeling is far from universal. Regardless of the limits of genetic determinism, there are thoughtful — not to say influential — people for whom full human cloning is a moral and emotional minefield that should place it among taboos such as cannibalism or incest.

Consent

Declarations at the national or international level have their place, particularly in strengthening the willingness of signatory governments to resist bullying and demagoguery, religious or otherwise. But the value of purely legislative action, as even many supporters admit, is limited. Laws will ultimately be effective only if they are rooted in the consent of individuals and if that consent is, in turn, based on a realistic assessment of both the scientific possibilities and social pitfalls of cloning.

Detailed exploration of both is needed. The first because the potential medical benefits are such that it would be verging on the immoral not to attempt to find out what, and how achievable, these might be — just as a proper understanding of the science behind biodiversity is important for an effective preservation strategy (see page 215). The second because it is essential (and not only for the health and credibility of the biotechnology industry) that society develops regulatory instruments sufficiently sensitive to distinguish between the responsible and the irresponsible application of whatever knowledge emerges.

There are signs that such exploration is beginning. Britain's Human Genetics Advisory Commission is due to produce a discussion paper within the next few weeks, said to propose the circumstances in which some forms of cloning-related research should be allowed to proceed. In the United States, similar debates are being held within professional medical societies (though so far there remains reluctance to grasp them firmly elsewhere, for example on the National Bioethics Advisory Committee). Human cloning is not the most pressing issue facing our health-care systems, or indeed, our religious or political leaders. But the way we handle it provides a measure of our maturity in addressing critical issues at the intersection between the research community and the society which has to confront the uncomfortable truths and capabilities that research can yield. □



France eyes role in US mission to retrieve rocks from Mars

[WASHINGTON & MUNICH] France is considering taking part in a proposed US spacecraft expedition to collect samples of Martian rock and return them to Earth. France's participation, if it comes about, could greatly increase the scientific return of the mission, which is planned for the year 2005.

The idea of France launching the 2005 spacecraft on Europe's large Ariane 5 launcher in exchange for a significant role in the US Mars exploration programme has been discussed for several months by officials at the French Centre National d'Études Spatiales (CNES) and the US National Aeronautics and Space Administration (NASA). NASA's rules prevent it from buying a non-US launcher directly.

No agreement has been reached, but the NASA Administrator, Daniel Goldin, said last week that he is optimistic that the sample return mission would be collaborative. "It is my hope and expectation that we will do this with the French," he told a meeting of the American Astronomical Society meeting in Washington, DC.

NASA intends to launch spacecraft to Mars at every available opportunity for the next decade or more, with launch windows occurring every 26 months. The current plan is to send landers equipped with rover vehicles to two different sites on the planet in 2001 and 2003.

Each rover would collect rock samples and 'cache' them for later retrieval. The 2005 lander would then be sent to the site deemed more scientifically interesting of the two,

where it would pick up the cached samples and return them to Earth in 2008.

Because of NASA budget constraints, the 2005 spacecraft is scheduled to launch on a medium-size Atlas or Delta rocket. But the Ariane, with its much larger capacity, would almost double the amount of payload that could be delivered to Mars. France could decide to use this excess capacity to send a lander to the second sampling site, according to Daniel McCleese, chief scientist for Mars exploration at NASA's Jet Propulsion Laboratory in California.

A full Ariane 5 launch would cost CNES about Ecu150 million (US\$160 million). Other, cheaper alternatives for France include providing the Mars-orbiting vehicle that receives samples from the surface.

The country could also add scientific instruments to the 2005 rover — currently designated to be a 'dumb' rover that simply fetches samples collected on earlier missions. Adding the ability to analyse and collect its own rocks would provide a backup in case the rover is unable to retrieve the cached samples, says McCleese.

Francis Rocard, who heads the Solar System exploration office at CNES, says that if France were to furnish an Ariane launch, it would want to be involved in "some key elements" of NASA's Mars programme in return. This would probably include participation in the 2001 and 2003 missions, as well as the 2005 sample return, he says.

NASA has already chosen the scientific payload for the 2001 mission, but French sci-



Capacity booster, Ariane 5 seen here at its first launch last October, would double the payload that could be sent to Mars.

entists could be added as co-investigators, says Rocard. The 2003 payload, which has yet to be selected, could also include French experiments.

CNES has asked Germany to join the collaboration should an agreement with NASA be signed. It is likely that German scientists would contribute to the instrumentation but Germany's space budget is currently too tight to make a large involvement likely.

CNES scientists have long had an interest in Mars. The agency sponsored several experiments on the Russian Mars 96 spacecraft that failed two years ago before reaching the planet. French investigators are especially strong in the areas of imaging spectroscopy and micro-seismometers for planetary surface studies, according to McCleese.

International participation would strengthen NASA's Mars programme, he says, and is of particular interest to Goldin. Working groups from CNES and JPL have been meeting regularly to identify technical opportunities for collaboration, while the French government is working out budgetary issues. Both McCleese and Rocard hope for agreement by this summer.

The discussions come at a time when the European Space Agency's own proposed Mars mission, called Mars Express, faces uncertain funding (see *Nature*, 390, 325;

French inquiry to look into thesis fraud

[PARIS] France's Atomic Energy Commission (CEA) and the Centre de la Recherche Scientifique (CNRS) have jointly set up a panel of inquiry to investigate the circumstances surrounding the apparent fabrication of data in a doctoral thesis by a young graduate student, and the 'structural factors' that initially allowed the fraud to take place.

The situation concerns a researcher who appears to have manipulated her results to make them correspond to the conclusions she was seeking. Although doubts had arisen over the quality of her data when she presented her thesis, she was still recruited by the CEA as a postdoctoral researcher.

Several months later, however, it became clear that some of the data had been fabricated. The researcher — who has not been named — was removed from her post,

and two publications in which the results appeared are now under review.

Catherine Cesarsky, director of the department of science of matter at the CEA, says that this is the first clear example of scientific fraud she has come across. She adds that it is "very rare" for students in her department to be left alone to work on data, as most work in large teams, making fraudulent activity virtually impossible.

But the CEA now faces the question of how the student was able to publish her results, obtained while she was working in a laboratory run jointly with the CNRS, and include them in her thesis. This will be one of the main questions to be addressed by the inquiry. "We want to reassure ourselves that a student is never again left with data that no-one else has checked," says Cesarsky. The panel will report within six weeks. **Eric Glover**



Rover returns: a vehicle would collect rock samples to be brought back to Earth.

1997). CNES has been encouraged by French research minister Claude Allègre — himself a geologist who took part in analysis of the moon rocks — to play a large and consistent role in the international Mars programme and to do so it is prepared to forge significant bilateral agreements. This is a situation that secretly worries the ESA considerably, since bilateral agreements inevitably compete for national funding with scientific instruments for ESA missions.

Mars Express, a lander/orbiter mission scheduled for launch in 2003, may also include international collaboration. ESA and NASA have been discussing an Italian idea whereby a communications link on the Mars Express spacecraft would be used for the US sample return mission in 2005. So far, however, no agreements have been signed.

CNES and NASA are farther along in discussing what McCleese calls joint “precursor mini-missions” to Mars, which would be piggybacked on Ariane 5 launches that deliver commercial satellites to high geostationary orbits. These missions, which would run in addition to NASA’s already planned Mars programme, could deliver between 80 and 160 kg of payload to the Martian surface beginning as early as 2000.

France would provide the Ariane 5 piggyback ride and an upper stage rocket, while NASA would provide the spacecraft. The scientific instruments would be developed jointly. Each mission would cost between \$25 million and \$50 million, and would be able to land simple instruments such as small probes or penetrators on the surface of Mars.

McCleese says the mini-missions would be ideal for positioning networks of sensors that could collect seismological and meteorological data. “It’s something we’ve been working on for 20 years,” he says.

US funding for these mini-missions would have to come out of NASA’s existing Mars programme, which is budgeted at about \$1 billion over 10 years. The missions might also be accommodated within the agency’s Discovery line of small and inexpensive planetary missions. As with collaboration on the sample return, no agreement has been reached yet. But McCleese believes the mini-missions have “a very high probability of happening”.

Tony Reichhardt

Quebec says yes to federal money for infrastructure

[MONTREAL] The separatist government of the Canadian province of Quebec has bent its sovereignty principles to gain access to some of the C\$800 million (US\$560 million) of research infrastructure funding being offered by the federal government.

But the move has not been made without protest. And it remains unclear how far the province is likely to go in imposing what it has described as its own priorities on the way the money is used.

The funds are being made available through the Canada Foundation for Innovation (CFI), which was set up last year in the federal budget using money released as a result of a tight fiscal policy (see *Nature* 385, 759; 1997). It was hoped that the funds would help to attract back to the country Canadian researchers working abroad.

Project proposals were invited from universities and hospitals across the country, but Quebec’s education minister, Pauline Marois, condemned this in December as an example of federal intrusion in a provincial jurisdiction.

Education and the administration of health services are provincial responsibilities in Canada (although these are not the only areas in which Quebec frequently complains about federal intrusion). Marois demanded that Quebec’s share of the C\$800 million be handed directly to the province, and even threatened to penalize institutions that might respond to the CFI’s invitation.

When the federal government saw the contradiction in Quebec policy, its industry minister, John Manley, sent a letter to Marois and Quebec’s health minister Jean Rochon accusing the province of playing petty politics.

At a news conference, Quebec’s intergovernmental affairs minister, Jacques Brassard, revealed his government’s underlying philosophy when he described the “ideal situation” as being able to say to the federal government: “It’s none of your business. We don’t want you in Quebec and we don’t want your money.”

But Brassard added that sometimes it is necessary to be pragmatic. A modified procedure has now been developed under which Quebec universities and hospitals will submit their projects to the CFI through the Quebec government.

One reason for Quebec’s stance is that — like the other provinces — it will have to pay overhead costs, and perhaps provide matching funds, for any projects funded by the CFI. For some years, all provinces have suffered drastic cuts in federal transfer funds in education and health fields.



Denis Gagnon, acting president and chief executive officer of CFI, says that, like many researchers, he was greatly disappointed by the initial stance of the Quebec government.

He says Quebec came to the press conference at which Brassard made his comments with both a political issue and an operational issue. “The political issue I have nothing to do with,” said Gagnon. “It’s between governments.”

He admits that all provinces are concerned about the question of operating costs, and will be meeting their representatives in the next few weeks to talk about the issue.

“I will be discussing with the provinces what we want to do and how we could work together in such a way that we will fund what they want.”

Gagnon thinks it unlikely that the provinces will seek to impose their own priorities on research institutions, and is keen to receive applications that have the support of these governments as well as their institutions. But he admits he does not know how this might come about.

For example, Quebec might ask its own two fund-granting institutions, or some sort of committee, to consider the applications before they go to the CFI.

Gagnon says it is unclear what the Quebec government meant when it said that applications should accord with “our priorities”.

“It’s clear from the original funding agreement between CFI and the federal government that we would like to have the universities and hospitals establish their [own] priorities, and then tell us what they are,” he said. “My feeling is that this is what will happen.”

David Spurgeon

Biodiversity body 'needs more science'

[LONDON] The United Nations' biological diversity convention needs to strengthen its scientific advisory body, delegates from member countries concluded at a meeting last week in London.

The meeting was held primarily to discuss the workings of the UN Convention on Biological Diversity, signed at the Earth Summit in Brazil in 1992. But delegates to the workshop spent much of their time considering the convention's Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA), focusing on the long-standing concern that the body's meetings tend to concentrate more on political issues than on science.

SBSTTA was set up to give scientific advice to member countries. Its remit also includes analysing national biodiversity inventories — the first batch of which are waiting to be read — assessing the state of biodiversity research, and commissioning new projects in response to specific requests.

But the body has met few of these objectives. It has met only three times since the signing of the convention in 1992. Discussions have taken on a political flavour because governments are frequently represented by diplomats instead of scientists. The last meeting of SBSTTA was dominated by a dispute about which UN convention should take precedence in formulating forestry policy (see *Nature* 390, 432; 1997).

It is often left to the convention secretariat to collect scientific advice for parties to the convention. "There is clearly a need for change," says Calestous Juma, executive secretary to the biodiversity convention who is based in Montreal. "There is not even a body that can synthesize the available scientific information."

Even the guide that has been produced — the comprehensive, 1,100-page *Global Biodiversity Assessment*, which involved 1,100 experts from 80 countries — was coordinated not by the biodiversity convention but by the UN Environment Programme.

At the workshop last week, the 35 participants brought together by the UK government considered several options.

One was to set up a parallel, but independent, body of scientific experts who would review the status of biodiversity research, commission new projects and provide relevant advice to parties to the convention.

This body would be similar to the UN's Intergovernmental Panel on Climate Change (IPCC) which, through its decade-long efforts, contributed critically to the adoption of the Kyoto Protocol on greenhouse gases last month.

But this idea was recognized as being the most costly of those considered, as well as being politically unworkable in the short



Out of focus: the convention is meant to boost biodiversity research, as here in Costa Rica.

term. The main reason for this is the reluctance of donor countries to invest in a new institution. Also, the Global Environment Facility, the UN's main environment funding mechanism, has spent most of its \$2bn initial budget.

Another option would be to make better use of informal scientific expertise through national and international academies of sciences, and through organizations such as DIVERSITAS, a group of biodiversity researchers drawn from six international scientific bodies that include Unesco and the International Council of Scientific Unions.

The lack of available structured scientific advice has meant that, five years on from the signing of the convention, little has been added to the sum of scientific knowledge on biodiversity, and a number of important questions remain.

To answer some of these, it is necessary to identify and classify the unknown 80 per cent of the Earth's species; to agree on a common method of classifying plants and animals; to begin to unravel the complex mechanisms

through which different species depend on each other for growth and survival; to find a consensus on the rate of biodiversity loss; and to carry out research into the environmental implications of genetically modified organisms.

Genetically modified organisms are the subject of a proposed protocol on biosafety due to be adopted at the end of this year. The protocol is controversial. Some observers have questioned the decision to include such organisms in an international convention on biological diversity.

Others challenge the wisdom of a deadline for a protocol when many aspects of the environmental impacts of genetically modified organisms remain unclear.

Meanwhile, the secretariat — in conjunction with member countries — has already begun to set research priorities and has agreed on a 10-year programme of scientific work in four main areas: marine biodiversity; the impact on biodiversity of agriculture; forests and inland waters; and research on the interdependence of species — known as the 'ecosystem approach'.

And DIVERSITAS will bring together 15 experts to explore the scientific questions stemming from each of the convention's articles for a meeting in March at the Autonomous University of Mexico (Unam).

Sir Ghilleen Prance, chairman of DIVERSITAS and director of the Royal Botanic Gardens at Kew in London, says he is keen to involve more scientists in the biodiversity convention. "Up till now the politics has dominated rather than the science," says Prance. We have a challenge on our hands to change things."

Ehsan Masood

South Africa sells research to its children

[CAPE TOWN] South Africa has declared 1998 the 'Year of Science and Technology'. The programme will be launched formally in Cape Town next month, when President Nelson Mandela will visit a school computer laboratory and receive e-mail messages sent by schoolchildren at other schools.

The main aim of the programme, according to member of parliament M. F. Cassim of the arts, culture, science and technology portfolio committee, will be "to reach out to every child in every corner of the land" to persuade them of both the excitement and the importance of science and technology.

This will be done through a series of 'focus weeks', staggered throughout the year, in each of the nine provinces. In many cases these weeks will be followed by additional activities in rural areas.

Foreign governments have agreed to support the programmes in four provinces. The Western Cape Province, which will host the first focus week, will be supported by France, and will include visits to research vessels launched in Cape Town harbour. Focus weeks in Mpumalanga, the Northern Province and the North West Province will be supported by Russia, the Netherlands and the United States respectively.

South Africa's eight science councils will be participating through a range of exhibitions, competitions, public lectures and open days. Each is responsible for a particular 'thrust', with the Foundation for Research Development, for example, promoting astronomy and space science, the Medical Research Council showcasing public health, and the Council for Geoscience responsible for archaeology and palaeontology.

Michael Cherry

Clinton pledges more money for science next year

[WASHINGTON] Science in the United States will enjoy healthy funding growth next year, following a decision by the Clinton administration to bow to pressure from scientific societies and their supporters in the Senate.

The administration has said it will ask for the money in its 1999 budget proposal, to be released on 2 February. Officials say the budget proposal will ask for increases of more than 7 per cent at the National Institutes of Health (NIH) and 9 per cent at the National Science Foundation. There will be smaller increases above inflation at other science funding agencies.

The increases are part of a \$10 billion investment package which the administration says is partially contingent on the ability of Congress to pass complex legislation to settle a major dispute between tobacco companies and state governments. But officials say that the package will be funded from other sources if the legislation fails, and science lobbyists are confident that money will be found for the increases in the event of such a failure.

Clinton's decision to boost science funding has come after a year of intensive lobbying by societies representing three million scientists and engineers, culminating in a December letter-writing campaign by members of the American Physical Society and the American Chemical Society.

Congress had expressed its own readiness to boost research spending. On 4 December, a letter from four key senators, including Pete Domenici (Republican, New Mexico), the chairman of the Senate Budget Committee, asked the administration to respond to their proposal to double research spending over ten years. The administration was able to make a positive response after budget projections released earlier this month enabled Clinton to announce that he would balance the entire budget in 1999 — three years earlier than anticipated.

The vanishing budget deficit puts Clinton in a good position to win approval for his investment plans. But the proposals for science spending are likely to have particularly strong support from both parties in the Congress. Lobbyists are optimistic that the increase for the NIH will be considerably larger than Clinton has proposed.

Last Saturday (10 January), Clinton devoted his weekly radio address to science and technology, pledging to support them in the budget and to expand on his plans for them in his State of the Union address at the end of this month. "Our nation was founded by men and women who firmly believed in the power of science to transform their world for the better," he said.

Colin Macilwain

Collapse of charity boosts call for tighter oversight

[LONDON] The government agency that registers Britain's charitable organizations has come under fire for licensing a medical research charity that took £250,000 (US\$403,000) from the public and spent 80 per cent of this sum on "administrative costs" without contributing a penny towards research. The charity went into liquidation last year.

The Charity Commission is being criticized by the Association of Medical Research Charities (AMRC) for its handling of the Transplant Research Foundation, whose motto was: "Don't take your organs to Heaven — Heaven knows we need them here."

The foundation, which stated that its goal was to "support and carry out research into organ transplants and aftercare for the public benefit", went into liquidation following a year-long investigation by the commission. The investigation was prompted by complaints from researchers about delays in processing their grant applications, as well as about the use of "unconventional" methods for assessing project proposals.

The investigation revealed that the foundation lacked a standard procedure, and access to trained scientific personnel, for assessing research applications. The foundation had received 90 applications for grants worth £90,000, but no funds were given out.

The case has important implications for the licensing of medical research charities. Diana Garnham, general secretary of the AMRC, says she was taken aback at the Charity Commission's failure to spot what appeared to be clear irregularities.

Garnham adds that the case also shows that the commission still fails to understand the need for medical research proposals to be independently refereed, despite the publication of guidelines to this effect in 1994, one year before the Transplant Research Foundation was registered.

The guidelines followed a controversy involving Britain's two largest cancer chari-

ties, the Cancer Research Campaign and the Imperial Cancer Research Fund. These charities had funded and publicized the work of a PhD student at the Bristol Cancer Help Centre whose research later turned out to be flawed and which lacked adequate supervision.

The Charity Commission published its guidelines for medical research charities after the so-called Bristol Inquiry (see *Nature* 369, 428; 1994). But Garnham says that, despite the guidelines, the commission has acknowledged to the AMRC that it lacks expertise in medical research funding matters.

She warns that this could erode confidence in medical charities. "If the Charity Commission can publish guidelines and not enforce them, where does that leave the public who donate money?"

A spokesman for the commission says the foundation's financial situation was "not as healthy as they thought it was". But he adds: "At the time we registered the fund it did meet our criteria." The spokesman, however, was unable to confirm whether the criteria included compliance with the guidelines.

The investigation was prompted by a complaint from Hannah Gould, professor of molecular biology and biophysics at King's College, London, who was concerned at the methods used to assess research applications. Each applicant had been invited by the foundation to become a 'reviewer' for other grant applications and was asked to score these — as well as their own application — on a scale of 1 to 25. Applicants were left free to devise their own criteria for scoring projects.

Gould says she assessed 50 applications. She was also asked to be one of the foundation's trustees. "It was all very silly," she says.

Numerous efforts to contact the foundation's director general Rick Holland this week were unsuccessful. The telephone and e-mail numbers on the foundation's headed notepaper were no longer in service, and Holland's private telephone was not listed in the UK telephone directory.

Ehsan Masood

Hubble sheds new light on Saturn's poles

[WASHINGTON] Spectacular details of the auroral curtains of ultraviolet light around Saturn's north and south poles have been observed for the first time by the Space Telescope Imaging Spectrograph on board the Hubble Space Telescope. The photograph (right), released last week, was taken in October, when Saturn was 1.3 billion km from Earth.

When used as a camera, the new instrument is more than ten times as



sensitive as previous Hubble instruments in the ultraviolet. Saturn's auroral displays are caused by an energetic wind from the Sun

that sweeps over the planet. Unlike Earth, Saturn's aurora is seen only in ultraviolet light invisible from the Earth's surface.

Controversial cancer drug wins local approval in Italy

[ROME] The southern Italian region of Puglia announced last week that it is prepared to provide the expensive drug somatostatin free to cancer patients, even though the drug has no proven efficacy in most cancers and is not on the national list of reimbursable drugs.

The decision represents the latest move in a four-month battle between the conventional medical community and the supporters of an unconventional approach to cancer therapy claimed to be based on stimulating the body's self-healing properties. The situation echoes the dispute over the controversial drug laetrile in the United States in the 1970s.

The so-called Di Bella approach was developed by an 85-year-old physician, Luigi Di Bella, from Modena near Bologna. He argues that established approaches to therapy directed at cancerous cells are fundamentally flawed, and that cures can be achieved only by stimulating the body's self-healing properties.

The battle over Di Bella's claims has all the ingredients of classic Italian high drama, and has come to a head in a direct conflict between the judiciary and the state. Several judges have ordered that the drug, which can cost up to US\$6,500 a month, should be made available to individual patients free of charge because of the guarantee in the Italian constitution to protect the health of individuals.

But Italy's minister of health, Rosa Bindi, insists that "the judiciary cannot decide on therapy", and has resisted an emotional campaign led by relatives of cancer patients and doctors who support Di Bella's unconventional therapy.

Di Bella tailors a cocktail of drugs and vitamins to individual patients, but the mix always includes somatostatin. He claims to have cured thousands of patients in this way over the past 30 years, although he has never published details of his results.

Although somatostatin is expensive, many patients who have not been cured by conventional chemotherapy have turned enthusiastically to the Di Bella method. So-called Di Bellists have organized frequent demonstrations to demand state support for the treatment, and many have taken their claims to court.

The story hit the headlines last month after a judge in Lecce, Puglia, ordered the state to pay for his plaintiffs' treatment. Several other local judges followed, and the media fanned the flames, portraying conventional medicine as a conspiracy condemning poorer cancer patients to die.

Before Christmas, Bindi ordered Di Bella to submit his files on patients to the health ministry by this week. But he has been reluctant to do so, contending that the medical establishment will reject his treatment method out of hand.

Di Bella's supporters have also reacted with hostility to an offer by Bindi to assemble an international panel of independent experts to help to assess the Di Bella files and help to decide whether there is a strong enough scientific basis to warrant a clinical trial.

A spokesman for the Italian Association for the Assistance of Neoplastic Patients says it has no faith in the fairness of an international committee. It is demanding that the ministry create a series of Di Bella cancer treatment centres where the results of the treatment could be monitored.

Parliament's social affairs committee meets this week to discuss what it describes as "the conflict of competence between judges and the governmental scientific

institutions". But the problem is so sensitive that it will not be resolved simply by reference to law and logic, according to Giuseppe Benegiano, director of the Higher Institute of Public Health in Rome.

"The issue would be totally emotional anywhere because it involves the dying," says Benegiano. "And since Italy is a country famous for its emotion, the effect gets multiplied." Given the "explosive mixture of issues," he says, the only solution may be to conduct a full clinical trial, even though available data do not provide strong support for such a move.

In the US case, laetrile extracted from apricot stones was hailed as a miracle cure for cancer in the 1970s. Despite the absence of scientific evidence of anticancer activity, several states defied the Food and Drug Administration and allowed the drug to be prescribed. Public pressure finally forced the agency to conduct trials, which proved negative.

Umberto Veronesi, scientific director of the European Institute of Oncology in Milan, admits to watching the Italian drama with resignation verging on despair: "We've come to expect this sort of claim [for a miracle cure] every two or three years. But eventually they get forgotten." **Alison Abbott**

Fire damages Germany's Einstein tower



[BERLIN] Germany's best-known piece of scientific architecture, the Einstein tower, was damaged by fire last week after a spark ignited solvent used in renovation work.

The building's scientific instruments remain intact, however, as they had been removed last autumn in preparation for the overhaul, which was needed following decades of neglect during communist rule in East Germany.

The tower (inset), which is in Potsdam near Berlin, was built in the expressionist style in the early 1920s by the renowned architect Erich Mendelsohn as a solar

observatory to test predictions of Einstein's general theory of relativity. It remains an important reference in the history of European architecture, and has continued to play a scientific role. For example, it made solar observations in parallel with the European Space Agency's space-based solar observatory SOHO, which was launched in December 1995.

Günther Hasinger, a director of the Astrophysical Institute in Potsdam which owns the tower, says that damage to the tower's infrastructure will be repaired by the end of the year. **A.A.**

Seed sows further doubts on cloning

Meredith Wadman

Plans by a Chicago embryologist to clone a human being have renewed pressure for swift action to ban the practice in the United States. But scientists warn that an ill-drafted law could stifle genuine research.

[WASHINGTON] The provocative pronouncements of Richard Seed of Chicago, a Harvard-trained physicist who last week announced plans to clone a human being, appear to bear little relationship to reality. Seed has little money, few if any fertility physicians willing to help him, and no laboratory.

Furthermore, prohibitive scientific obstacles remain to such a feat. It took the Scottish scientist Ian Wilmut and his team 277 attempts to produce Dolly, the clone of an adult ewe (see *Nature* 385, 810; 1997), and nothing suggests that the effort in humans is going to prove any easier.

"This has been played on front pages as though there's something to announce," says Alexander Morgan Capron, a professor of law and medicine at the University of Southern California who is a member of the National Bioethics Advisory Commission (NBAC). "My own sense is that it belongs in the entertainment section."

But, however dubious its origins, the tempest set off by Seed's announcement points to the fact that, at least in the United States, cloning a human being using private funds would break no law.

That could soon change. Seed "will not

do human cloning in this country," declared Donna Shalala, Secretary of Health and Human Services, on CBS television last Sunday (11 January). And President Bill Clinton, in his weekly radio address last Saturday, called Seed's announcement "profoundly troubling" and urged Congress to enact legislation when it reconvenes this month. "There is virtually unanimous consensus in the scientific and medical communities that attempting to use these cloning techniques to clone a human is untested and unsafe and morally unacceptable," Clinton said.

He added that Seed's plans make it "clearer than ever" that cloning legislation he sent to the Congress last June is "exactly what is needed" (see *Nature* 387, 644; 1997). Clinton's Cloning Prohibition Act would ban for at least five years the cloning of human beings in the public and private sectors.

So far Congress has not acted on it. But the mood of legislators appears to be rapidly changing. Speaking on US television last Sunday, the House of Representatives majority leader, Richard Armey, declared Seed's project a "nasty business". Referring to a House bill already introduced by congressman Vern Ehlers (Republican, Michigan),

US ban 'would resist challenges in court'

[WASHINGTON] The increasing likelihood that the US Congress will pass legislation outlawing human cloning immediately raises constitutional questions. Would such a ban withstand court challenges that it breaches the freedom of scientific enquiry, for example? And would it violate a right to reproductive freedom?

Constitutional scholars interviewed last week suggest that, if a cloning law were challenged and the case reached the US Supreme Court, arguments for neither procreative nor scientific liberty would hold sway.

According to Richard Fallon, a professor of constitutional law at the Harvard Law School, with human health so evidently at risk, the right to freedom of scientific enquiry, which has some basis in lower court decisions, would appear "frivolous".

Neither would the more powerful argument ultimately triumph that couples should have a right to reproduce without government interference, say legal scholars.

"What the court would be likely to say is that there's a difference between reproduction and replication," says Cass Sunstein, a

professor of law at the University of Chicago.

Sunstein suggests that the court would probably rule that earlier decisions establishing some rights of reproductive freedom do "not extend to an unfettered right to reproduce by whatever technology science comes up with".

Even if such a right were to be acknowledged, he and others suggest that the state would have sufficient ammunition to argue convincingly that that right should be outweighed by the need to protect developing fetuses and resulting babies from any undue physical harm.

M. W.



Out of step: Seed faces opprobrium from the scientific community as well as legal hurdles.

Armey said: "We are going to move that ban." And Senator Bill Frist (Republican, Tennessee), the cardiac surgeon who chairs the Senate subcommittee that has jurisdiction on the issue, told *Nature* through a spokeswoman that he is considering introducing legislation to this end. "If Richard Seed gets credit for anything, it's for putting this issue back before Congress," says Capron.

The lack of action in Congress has occurred despite a flurry of hearings that followed the announcement last February of Dolly's birth, the recommendation last June by NBAC that Congress should outlaw human cloning for a period (see *Nature* 387, 217; 1997), and the submission to Congress of Clinton's bill, which in essence codifies the commission's recommendation.

Three bills authored by congressmen have also languished — two of them would bar federal funding for research on cloning humans and one would impose civil fines on anyone who produces a human clone (see *Nature* 386, 97; 1997).

Before Seed's announcement, Frist had been reluctant to legislate. The chairman of the public health subcommittee of the Senate Labor and Human Resources Committee said in October that "there is no reason for Congress to legislate" and that the issue ought instead to be debated in a "national dialogue".

According to Margaret Camp, a spokeswoman for Frist, the senator thought until last week that a voluntary private sector moratorium on cloning called for by Clinton in March was being followed. That, com-

bined with a ban on federal funding for human cloning imposed by Clinton at the same time, was considered by the senator to be sufficient. But, says Camp, Frist is now reconsidering his position in the light of Seed's announcement. "He is looking at legislation now. But it's with human cloning [for reproductive purposes]. He feels strongly about protecting research."

Behind the congressional inaction on cloning over the last ten months has been not only a busy schedule in which the issue got pushed aside, but also abortion politics, as exemplified in the fate of the Clinton bill (see *Nature* **387**, 748; 1997).

Despite courting Republicans and Democrats in the House and Senate, and support for the bill's approach from groups such as the Biotechnology Industry Organization, the White House has been unable to find a sponsor for the bill. This is partly because the bill — like the recommendations of the NBAC — does not prohibit use of cloning techniques in human embryo research.

Even Connie Morella, the liberal Republican congresswoman who represents the Maryland constituency that includes the National Institutes of Health, has not stepped forward to sponsor the legislation — although the White House courted her. Morella called last week for a bill to ban human cloning.

"Nobody in Congress wanted to touch [the Clinton bill] with a ten-foot pole," says Roger Pedersen, a mammalian embryologist at the University of California, San Francisco, who speaks for the Federation of American Societies for Experimental Biology (FASEB). The federation, which represents 52,000 sci-

Europe brings in first international ban

[PARIS] At a ceremony in Paris on Monday (12 January), 17 European countries signed a protocol added to the European Convention on Human Rights and Biomedicine that bans the use of human cloning for reproductive purposes — the first legally binding international agreement to do so.

Two countries were conspicuous by their absence among the signatories: the United

Kingdom and Germany. Neither can sign the protocol because they have not signed the convention itself, the United Kingdom because of delays caused by the change in government, and Germany because it feels the provisions in the convention concerning human embryo research and consent by incapacitated individuals are not strict enough.

But British legislation setting up the Human Fertilization and Embryo

Authority already forbids the use of human cloning, and human cloning for reproductive purposes is banned by law in Germany.

The signing of the agreement was welcomed by Jacques Chirac, the French president. Speaking at a meeting of Europe's ethics committees in Paris, Chirac said that an international ban was essential as otherwise the technology would migrate to countries where regulation was less strict. **Declan Butler**

entists, is pushing a different approach. In September, following the lead of the Society for Developmental Biology, it adopted a voluntary moratorium on cloning.

Since then, the American Society for Reproductive Medicine, representing 9,000 fertility physicians, and the Society for the Study of Reproduction, with 2,400 members, have adopted the same moratorium. Pedersen says the opprobrium heaped on Seed by fellow scientists last week is evidence of the power of peer pressure in the face of aberrant behaviour. "There's a circling of the wagons, with him outside," he says.

Although FASEB does not condone human cloning at this stage, he adds, "Congress has been acting responsibly by not rushing legislation. Hastily written law can do more harm than good by deterring poten-

tially beneficial research. That's why I think a voluntary moratorium is more effective."

But Capron of the NBAC says that congressional inaction has resulted in domestic and international US failures. Domestically, he says, it has allowed a "cowboy" like Seed to pursue his ends unchallenged. And internationally, the fact that the United States has not condemned human cloning officially by passing the Clinton legislation "means that we're not in a leadership position".

The Clinton legislation would ban the use of cloning to create children by anyone in the private or public sector. It would expire after five years, before which time the NBAC would be required to review developments and recommend whether the law should be extended. It includes explicit protections for cloning's research applications in DNA, cells, tissues and animals, but does not mention human embryo research. Embryo research — whether or not using cloning technology — is at present legal if financed privately, but barred from federal funding.

Scientific groups have said that they fear that, unlike the Clinton bill, those introduced by congressmen are so broadly worded and imprecise that they threaten research. A Senate bill introduced last February by Christopher Bond (Republican, Missouri) would bar the use of federal funds for "research with respect to the cloning of a human individual". Bond said last week that he would push for a broader "emergency ban" to encompass the private sector.

Two House bills were introduced last winter by Vern Ehlers. One banned federal funding for "research that involves the use of a human somatic cell for the process of producing a human clone". The other imposed a penalty of up to \$5,000 on anyone "producing a human clone". Ehlers later expanded the first bill to include a ban on use of cloning in human embryo research. The House Science Committee approved it last July; it had to be rewritten to win enough votes to be passed by the committee. □

Unesco declaration lacks legal teeth

[PARIS] Richard Seed's declared intention to set up a cloning laboratory in Tijuana, Mexico, if the US Congress bans human cloning was quickly rebutted by the Mexican government, which last Friday expressed its "broadest repudiation" of Seed's proposal.

But, like Seed's challenge to the US Congress, his offshore plans point out another reality: the international community has no solid legal front to present against the planned work of Seed or those who may follow him. Not only could Seed now proceed in Mexico — where the president of the national bioethics commission last week could only implore the Mexican

Congress to legislate on the matter — but no binding international ban on cloning is now in place.

The 186 member states of the United Nations Educational, Scientific and Cultural Organization unanimously passed a declaration in November calling for a cloning ban (see *Nature* **390**, 110 & **388**, 508; 1997). But that declaration is not legally binding.

Noëlle Lenoir, the chairwoman of Unesco's International Bioethics Committee, says the committee cannot take a position on the need for a legally-binding international agreement, as it has frozen its activities pending a review of its functions to fulfil the

new mandate given to it by Unesco of following up the implementation of the provisions of the declaration. The new committee, which should be in place next month, will probably consider whether a legal instrument is desirable.

Turning the call for a ban on cloning for reproductive purposes contained in the declaration into a legally-binding agreement may in practice change little, argues Lenoir. What is important is that countries translate international agreements into their national laws and enforce them. "Pressure for international agreements counts for little unless there is national government will to implement them." **M.W. & D.B.**

Go-ahead given for trials of HIV vaccine in human volunteers

[WASHINGTON] A Californian biotechnology company has won approval from the US Food and Drug Administration to carry out the first large-scale trial of the controversial gp120 HIV vaccine in humans. The company, VaxGen, based in south San Francisco and headed by Donald Francis, a pioneer of AIDS research, will begin this year a three-year trial involving 7,500 healthy volunteers.

The \$20 million, phase-three trial — the first for an HIV vaccine — will test the vaccine's clinical efficacy in volunteers in the United States and Thailand. Critics had earlier challenged the efficacy of gp120, which relies for immunogenicity on a single subunit protein of HIV. In 1994, an advisory panel to the National Institutes of Health recommended against proceeding to phase-three trials, which establish a vaccine's clinical effectiveness in large numbers of volunteers.

John Moore, an AIDS researcher at the Aaron Diamond AIDS Research Center in New York City, describes the trial as a “total waste of time and money”. But Francis says there is “very broad support” for the trial. “We followed a very logical path [and] all the

data point to a positive result,” he says, adding that the \$20 million for the trial has come mainly from private sources. Genentech, Inc. has a 25 per cent stake in VaxGen.

Fukui, father of ‘frontier electronic theory’, dies

[TOKYO] Kenichi Fukui, Japan's only winner of a Nobel prize for chemistry, died of cancer last week in Kyoto at the age of 79. Director of the Institute for Fundamental Chemistry and professor emeritus at Kyoto University, Fukui was well known for his ‘frontier electronic theory’, which he published in 1952.

The theory, now applied widely in fields ranging from drug development to the conversion of solar energy, provides a universal principle for organized chemical reactions. Fukui shared the Nobel prize for chemistry in 1981 with Roald Hoffmann from Cornell University in the United States. Fukui was also known for his efforts to promote science education in Japan.

New England fishery edges back from brink

[LONDON] The collapsed New England fishery is beginning to show signs of recovery, even though stocks still remain dangerously low, according to a report by a committee of the US National Research Council (NRC). “Little

by little, existing management measures appear to be working to decrease fishing mortality in four of the five stocks reviewed,” it says in a report published in Washington DC last week.

The committee concludes that strict management measures must remain if the recovery is to continue. The NRC was asked by Congress last year to review US and Canadian fish stock assessment methods. Committee members also recommend that more independent fisheries scientists should be involved in peer-reviewing federal fish stock assessments.

Fossil fuels drove up the heat during 1997

[LONDON] Human activities, such as the burning of fossil fuels, contributed to 1997 being the hottest year on record, according to scientists at the US National Oceanic and Atmospheric Administration (NOAA). This is the first time that NOAA researchers have attributed temperature rise to human activities; in the past, they have reported their findings without speculating on the cause.

According to Tom Karl, a senior researcher at NOAA, the Earth's average temperature during 1997 was 16.92 °C. This is 0.42 degrees hotter than the average for the years 1961 to 1990. The average temperature in the previous hottest year, 1990, was 16.83 °C.

Pendergast quits FDA for job in industry

[WASHINGTON] A senior staff member once tipped as a potential head of the US Food and Drug Administration (FDA) has left the agency. Mary Pendergast, a deputy commissioner and senior adviser to the acting FDA chairman, left the agency on 2 January to establish a Washington lobbying office for Elan, a biopharmaceutical company based in Dublin, Ireland.

Pendergast could not be contacted last week, but Carl Feldbaum, the president of the Biotechnology Organization, called her “extremely responsible, knowledgeable [and] a tough negotiator”. Pendergast began as a lawyer for the FDA, and was promoted by former commissioner David Kessler, carving out a role for herself as the agency’s crisis-manager.

Russian academy stands firm against institute cuts

[MOSCOW] The Russian Academy of Sciences has decided that 98 per cent of its institutes will be included in a document being presented to the government for official accreditation as ‘scientific’ bodies. The decision has been taken despite the government’s declared intention to reduce significantly the number of such institutions,

and its assumption that this move would lead to a reduction of at least one-third in the number of bodies permitted to use the title.

At the same time, Andrei Gonchar, the vice-president of the academy, announced that the salaries of researchers in its institutes are to be increased by between US\$140 and US\$200 per month, depending on rank. Meeting this goal on current funding plans will require a 10 per cent reduction in staff, apparently in contradiction of the decision to leave staffing levels almost unchanged.

Researchers told to follow winning formula

[LONDON] Scientific organizations should consider basing their strategy on that of the low-price but highly successful French hotel chain *Formule 1* — which concentrates on providing factors that clients find most important, such as hygiene and a comfortable bed, and virtually ignoring the rest, including eating facilities — according to John Krebs, the chief executive of Britain’s Natural Environment Research Council.

Krebs bases his conclusions on an analysis of *Formule 1*’s winning strategy published last year in the *Harvard Business Review* by two researchers, W. Chané Kim and Renée Mauborge. “It may seem a long way from motorway hotels to science,” writes Krebs in the most recent issue of the research council’s

magazine, *NERC News*. “But the concept of analysing key success factors and determining which are the ones to really focus on surely has general applicability.”

Shoemaker’s ashes fly with Lunar Prospector

[WASHINGTON] A suite of scientific instruments and a small capsule containing the ashes of the lunar geologist Eugene Shoemaker were on board the US Lunar Prospector spacecraft as it began orbiting the Moon last Sunday (11 January). The purpose of the \$63-million year-long mission is to produce chemical, magnetic and gravitational maps of the Moon and to look for evidence of water in the form of ice. It will also search for signs of radioactive gases such as radon leaking from the lunar interior.

Shoemaker, who died in an accident in Australia earlier this year, was among those who argued that a radar instrument on board the 1994 Clementine spacecraft found evidence of trapped ice in a permanently shaded region near the lunar south pole. Observers using the Arecibo radio telescope have said the radar signals are probably caused by surface roughness rather than by ice deposits. One of Prospector’s main objectives is to settle the controversy, perhaps within its first month of operation.

Impact takes precedence over interest

Sir — Johannes Stegmann¹ presents an apparently simple method to calculate (not “evaluate”) Journal impact factors (JIFs) for journals not receiving an official JIF through the *Journal Citation Reports* (JCR) of the Institute for Scientific Information (ISI). Unfortunately, there are some technical limitations to its applicability.

ISI's citation indexes include references to all kinds of bibliographical materials (also books and low-profile journals). An essential condition, however, is that these are cited by a controlled set of source journals. This implies that self-citations of non-source journals will not be included in the citation indexes. Although there is a lot of variation between individual journals, a substantial part of citations received are generally self-citations. In many cases the journals are their own single biggest source of citations.

Whereas for top-ranking journals receiving tens of thousands of citations the percentage of self-citations generally remains low, for a large part of the other journals a self-citation rate between 10% and 25% appears to be typical, especially when relating to the more recent years on which JIF calculations are based. So the total number of citations for constructing JIFs would be underrated to some degree. For example, deleting self-citations from the totals of the six source journal examples given by Stegmann would decrease their

impact factors for 1996 by a minimal 1.7% (for *Molecular Medicine* which, having started in 1994, can be considered a special case) to a more substantial 12.9% (for *International Journal of Developmental Biology*). More impressive examples are easily found (for example, *Molecular and Biochemical Parasitology* 21.3% or *International Journal of Leprosy* 30%).

Calculating the number of source items per journal may also be more difficult than Stegmann suggests. If the exact number of source articles of a specific journal is included in the SciSearch database, then it is probably already an ISI source journal featuring a JIF. Databases such as Medline do not necessarily cover all their source journals completely (especially if their contents are multidisciplinary, as for example with *Nature*). And, as acknowledged by Stegmann, the criteria for counting source articles may vary in different databases. Besides, lots of journals are not represented at all in the major databases. So having physical access to these journals would often be a requisite.

Assuming that JIFs are appropriate value indicators for scientific publications², one may argue that an approximate JIF is preferable to nothing, even if both numerator and denominator are inaccurate. Apart from this issue, one should bear in mind that the somewhat inaptly named JIF indicates the average

value of individual research articles judged by the journal they are published in, rather than the total impact of that journal. It would seem obvious that, when two journals have an identical JIF, if one annually publishes 2,000 papers and the other a mere 20, their overall impact on the scientific community cannot seriously be considered equal.

On the other hand, an important drawback of the unreserved attention being given to citation counts nowadays lies in the enormous gap between popular research areas (with many thousands of authors, papers and citations) and less popular research areas (with far fewer authors, papers and citations). To improve their status (and funding), the danger seems real that scientists would be tempted to neglect or abandon altogether the less popular research topics to the advantage of more rewarding ones. That might lead to an impoverishment of scientific knowledge in lots of domains.

Dirk Schoonbaert
Gilbert Roelants

*The Library,
Prince Leopold Institute
of Tropical Medicine,
Nationalestraat 155,
B-2000 Antwerpen, Belgium
e-mail: bib@itg.be*

1. Stegmann, G. *Nature* 390, 550 (1997).

2. Schoonbaert, D. & Roelants, G. *Trop. Med. Int. Health* 1, 739–752 (1996).

Challenges for vaccine institute

Sir — David Swinbanks¹ has succinctly outlined the political, fiscal, commercial and technical challenges faced by the Seoul-based International Vaccine Institute (IVI). Pending completion of the building of its laboratories, IVI is planning epidemiological studies of the burden of vaccine-preventable diseases, an assessment of vaccine requirements and clinical trials and field studies.

IVI must also concern itself with problems in the field for vaccines, therapeutics and diagnostics in Asia. There are alarming reports of large-scale theft and resale on the black market of expired and potentially toxic vaccines in Myanmar (formerly Burma). Nearly 10,000 doses of unrefrigerated vaccines were on sale in 1995. The vaccines were many months past the printed expiry date, had lost their potency and were toxic when injected into laboratory animals². Moreover, the use of expired drugs is an established practice in

many countries, such as Sierra Leone, that have no organized system for monitoring reaction to drugs³. Following the use of pre-tested blood for HIV with expired or improperly stored antibody screen reagents, the risk of HIV transmission in Zambia was at least six times as great as expected⁴.

Such events would be the rule rather than the exception in the event of global warming because of climate change or the effects of El Niño. IVI should start field studies immediately to monitor different permutations of temperature, humidity, atmospheric pressure and air velocity in remote parts of Asia. These parameters should be monitored with appliances such as electronic loggers. Monitoring of vaccine storage temperature by such loggers in Adelaide, South Australia, showed inadvertent vaccine exposure to sub-zero temperatures as well as to temperatures of more than 22 °C (ref. 5).

The data compiled by IVI would be of immense value both to the institute itself and to its partners in the venture. For instance, the World Health Organization would be able to modify the established

criteria for stability of vaccines, prophylactic and therapeutic, to ensure intact potency even in black-market sales of products past their use-by date. The manufacturers, in the West and elsewhere, could select products stabilized against the adverse environment described by IVI. Last but not least, the risk of an iatrogenic HIV spread during blood transfusion would be minimized through the availability of sturdy antibody assay kits. The frequent field exposure of IVI personnel before their premises are ready could significantly reduce the hostility of national/international manufacturers and funding agencies.

Subhash C. Arya
*Centre for Logistical
Research and Innovation,
M-122 (of part 2),
Greater Kailash-II,
New Delhi 110048, India*

1. Swinbanks, D. *Nature* 389, 655 (1997).

2. Masood, E. *Nature* 374, 669 (1995).

3. Sesay, M. M. *Int. Pharm. J.* 8, 202–206 (1994).

4. Costen, E. C. J. et al. *Transfusion* 37, 930–934 (1997).

5. Wawryk, A., Mavromatis, C. & Gold, M. *Br. Med. J.* 315, 518 (1997).

The eyes have it

Laurence D. Hurst and Andrew Pomiankowski

Females of many species prefer mates with extravagant traits. Studies of stalk-eyed flies show that, in this case at least, such preference is linked to suppression of a selfish gene that influences the sex ratio of offspring.

Peacocks would not have such large and elaborate tails were it not for the fact that peahens prefer their partners to be so ornate. But why are females so choosy? There are two standard answers. Fisher¹ suggested that males have elaborate traits purely because females find them attractive. In this case females must be choosy to gain the advantage of passing on the attractive trait to their sons, and so improve the chances of passing on their genes to subsequent generations. Others have conjectured that showy traits indicate the quality of a male². If only peacocks with 'good genes' can afford to bear bigger tails, peahens should mate with them to ensure that their offspring are viable and healthy.

Work by Gerald Wilkinson and colleagues³ on stalk-eyed flies (Diopsidae), described on page 276, provides the strongest evidence to date about the nature of some of the genes females prefer. As their name suggests, stalk-eyed flies have their eyes perched on the end of side-projecting stalks (see the cover of this issue). These can reach ridiculous proportions. In two of the species studied, males have considerably greater eye span than females, and male eye span even exceeds body length. In both of these species females show a strong preference for males with a large eye span. In a third species, male eye span was

much shorter (indeed no different to female eye span) and there was no female choice.

Wilkinson's group noticed that, curiously, the two species in which females preferred a large eye span also tended to have populations with more females than males. Closer examination revealed that some, but not all, broods were almost exclusively female. In contrast, the one species without female preference had an even sex ratio and no evidence that individual broods were female biased.

How does the sex-ratio bias come about? Some of the males, it transpires, have an unusual X chromosome⁴. At the time when sperm are produced, by the process of meiosis, gene(s) on this X chromosome somehow destroy most of the Y-bearing sperm. This is an example of X-linked 'meiotic drive' and has been described in several other flies (see box)⁵⁻⁷. Most of the progeny of males with the driving X are female as they inherit an X chromosome from both parents (Fig. 1, overleaf). This accounts for the biased population sex ratios.

From the X chromosome's point of view, drive is a profitable strategy, as more eggs end up being fertilized by sperm bearing the driving X. But the death of Y-bearing sperm is clearly bad for the Y chromosome. It is no surprise to discover that some males have Y-linked genes that can inhibit the activity of

the selfish X chromosome. The Y-suppressor can somehow reverse the direction of drive, leading to male-biased broods.

Could coexistence of female choice and the driving X chromosome be more than just a coincidence? To test this possibility, Wilkinson and colleagues subjected flies to 22 generations of artificial selection on eye span. Remarkably, they found that male eye span covaried with the ability to suppress the selfish X chromosome: males artificially selected for large eye span produced male-biased sex ratios; those selected for short eye span produced female-biased sex ratios. So females choosing males on the basis of their eye span will likewise be selecting a partner's ability to suppress the meiotic-drive gene. The covariance is consistent with the idea that at least some of the genes for long eye span are on the same Y chromosome as the suppressors of drive. Additional breeding experiments (G. Wilkinson, personal communication) reinforce this view.

Why should this linkage come about, and why do females prefer males with a large eye span? In the wild, the driving X chromosome coexists with the Y-suppressor, but not all X chromosomes drive and not all Y chromosomes suppress. Assuming that the frequencies of the four chromosomal types have reached equilibrium, there is by definition no selection in favour of the X-driver or Y-suppressor. However, at equilibrium the population sex-ratio is female biased and so, on average, each male has more offspring than each female. This means that selection favours mothers producing more sons than daughters as they will then have more grandchildren.

Now consider a rare autosomal gene (that is, one not on the X or Y chromosome) acting in females that codes for a preference for males with greater eye span. Will this gene

Selfish genes and meiotic drive

If one asks why an organism has a particular feature, such as eyes, it is usually supposed that it is because the trait is 'good for the individual'. This is another way of saying that although the gene for the feature must have been rare initially, it rose in frequency in the population because it increased the chances of the individual to survive and reproduce. 'Selfish genes' are not like this, for they are harmful to the individual. Unlike ordinary deleterious genes, however, selfish genes have some trick that allows them to increase in frequency when rare, despite

being harmful.

Meiotic drive genes are one class of selfish gene. The best-described example is *Segregation Distorter* in the fruitfly *Drosophila melanogaster*. This is autosomal (that is, not on the X or Y chromosome) and acts in males. Males that have the driver on one chromosome but not the other have half of their sperm killed, the half that do not contain the drive gene. This, of course, is not 'good for the individual'. But it is 'good for the gene', and the drive gene spreads in the population because of the increased numbers of eggs

that come to be fertilized by sperm containing it. Males in which both chromosomes have *Segregation Distorter* are sterile. As the driver spreads, this comes to be an increasingly common occurrence. In consequence, *Segregation Distorter* does not eliminate the original non-driving chromosome.

Although in no case is the mechanism of drive well understood, in all well-described instances the drive 'gene' is actually composed of two tightly linked types of gene. The simplest model proposes that one of these codes for a toxin. Even if only half of the

sperm have this gene, all sperm are affected by its toxic product – unless, that is, they contain the other gene, which codes for the antidote. Unlike the toxin, the antidote's activity is restricted to the sperm that contain the antidote gene.

Initially, most chromosomes will have neither the toxin nor the antidote. When rare, then, the two-gene complex is most likely to be in an individual whose other chromosome is sensitive to the toxin. This male will have half of its sperm killed and the driving gene-complex will spread.

L.D.H.

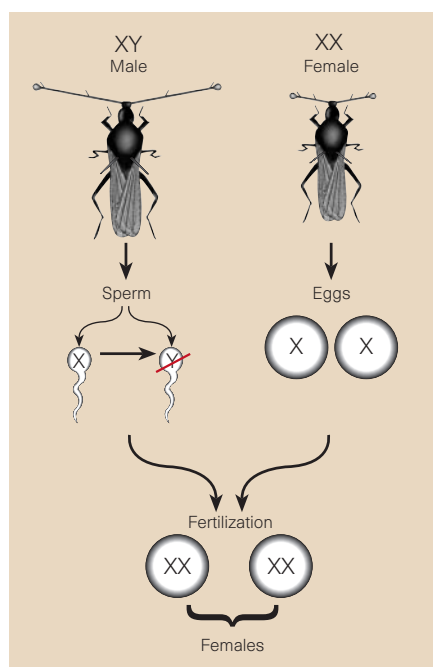


Figure 1 Meiotic drive in stalk-eyed flies. A male with a non-driving version of the X chromosome produces as many X-bearing sperm as Y-bearing sperm. On mating with a female, whose eggs are all X-bearing, half the progeny are female (XX) and half male (XY). However, if (as shown here) a male has the driving X, then, by an unknown mechanism, this X kills most of the sperm from the same male that contain the Y chromosome. Typically, males produce many more sperm than are necessary to fertilize the eggs of a female. So the eggs are all fertilized and nearly all inherit the father's driving X chromosome. As a consequence, the progeny are mostly female. If a male has the X-driver and the Y-suppressor, then the action of the X-driver is inhibited and the sex ratio even ends up a little male biased (about 63% male). In the absence of the driving X, 50:50 sex ratios are found, regardless of the type of Y chromosome.

become more common or will natural selection eliminate it? Spread of the gene depends on the linkage, in males, between genes influencing eye span and those for drive suppression. Preference will be favoured if genes for large eye span are on the suppressing form of the Y chromosome because males with the Y-suppressor produce male-biased broods. In contrast, preference for a large eye span is deleterious if genes for big eyes are on the non-suppressing Y chromosome, because males with this chromosome produce female-biased broods. Finally, preference will be neither favoured nor selected against if eye-span genes are located on autosomes, as these eye-span genes have no influence over the sex of offspring. So we can understand both when female preference will evolve and why at least some of the genes for the preferred trait are linked to Y-suppressors of meiotic drive. This model, however, cannot be the whole truth. For one thing, the

daughters of males selected for short eyes themselves prefer short-eyed males⁸.

Wilkinson and colleagues suggest another reason why female preference is favoured. In choosing males resistant to drive, the female is ensuring her sons have high fertility, because males with the Y-suppressor never have half their sperm killed. But this is only half of the equation. The Y-suppressor is probably costly when with a non-driving X chromosome (which would explain why some of the Y chromosomes do not suppress the driving X, and why at equilibrium there is a female-biased sex ratio). So the sons of a choosy female may do well compared to the average male with the driving X, but they do badly when compared to the average male without the driver. These forces will tend to balance out, in which case the effects of male fertility will be relatively unimportant.

Most people will find it surprising that the 'good genes' females prefer are not for general viability but for resistance to a selfish gene (see box). Few of those working on sexual selection would have imagined this answer, and few working on meiotic drive would have thought that their work related to the problem of female choice. That said, there is some suggestive evidence from mice⁹ and flies¹⁰ showing that females prefer males without meiotic drive (although in neither case was this preference associated with an extravagant trait, and there are counterexamples¹¹).

The fact that the genes appear not to be ones for general viability also upsets our understanding of why females usually prefer seemingly unnecessarily large male traits. Those espousing the 'good genes' argument suppose that such traits are costly and these costs ensure that trait size reflects male quality¹². In stalk-eyed flies, however, females do not seem to be assessing general male quality. In principle, the Y-linked suppressor is

equally likely to be in a poor-quality male as in a good-quality one. Nonetheless, the traits are large rather than small. So why is size important?

It is also unclear just how general an explanation of female choice this might be. Data^{5,6} showing that X-linked meiotic drive is more common than previously supposed support the idea that such preferences might be common, at least in flies. And in guppies there is one report of Y-linked variation for sex ratio¹³. (Is it a coincidence that guppies have numerous Y-linked sexually selected traits¹⁴?) But no good evidence of a sex-ratio distorter has been found in birds in which female choice is common. It would nonetheless be worth looking for a sex-ratio effect following artificial selection for trait size. □

Laurence D. Hurst is in the Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.

e-mail: l.d.hurst@bath.ac.uk

Andrew Pomiankowski is in the Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, UK.

e-mail: ucbbpom@ucl.ac.uk

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Williams, G. C. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, 1966).
3. Wilkinson, G. S., Presgraves, D. C. & Crymes, L. *Nature* **391**, 276–279 (1998).
4. Presgraves, D. C., Severance, E. & Wilkinson, G. S. *Genetics* **147**, 1169–1180 (1997).
5. Jaenike, J. *Am. Nat.* **148**, 237–254 (1996).
6. Merçot, H., Atlan, A., Jacques, M. & Montchamp-Moreau, C. *J. Evol. Biol.* **8**, 283–300 (1995).
7. Hurst, L. D. *Genetics* **142**, 641–643 (1996).
8. Wilkinson, G. S. & Reillo, P. R. *Proc. R. Soc. Lond. B* **255**, 1–6 (1994).
9. Lenington, S. *Adv. Study Behav.* **20**, 51–86 (1991).
10. Wu, C.-I. *Genetics* **105**, 651–679 (1983).
11. James, A. C. & Jaenike, J. *Anim. Behav.* **44**, 168–170 (1992).
12. Iwasa, Y., Pomiankowski, A. & Nee, S. *Evolution* **45**, 1431–1442 (1991).
13. Farr, J. A. *Heredity* **47**, 237–248 (1981).
14. Winge, O. *J. Genet.* **18**, 1–43 (1927).

Biological oceanography

Diatoms and the silicate factor

Victor Smetacek

Over most of the ocean, the amount of new production (net increase in organic matter) by phytoplankton is regulated by the availability in the surface layer of the plant nutrient nitrate. In the so-called 'high-nitrate, low-chlorophyll' (HNLC) regions, however, phytoplankton biomass remains low throughout the year despite high concentrations of nitrate. The condition has long been regarded as a paradox, but increasing evidence has indicated that, in these regions, phytoplankton production is limited by iron availability.

In their paper on page 270 of this issue¹, Dugdale and Wilkerson now add an additional facet to the debate. They analyse data from the equatorial upwelling zone (EUZ) of

the eastern Pacific — an HNLC region — and, with the aid of a simple budgetary model, show that the magnitude of new production there is a function of the input of dissolved silica (silicate) by upwelling of deeper water. This limitation occurs because new production is monopolized by diatoms (silica-shelled, unicellular algae) whose growth rates are limited by the supply of silicate.

The finding puts an upper limit on the rate of new production in the EUZ, and hence on the export of organic carbon from surface to deeper layers through the biological pump². This has implications for CO₂ exchange between ocean and atmosphere because the EUZ, a broad belt of water stretching across the entire eastern Pacific,

annually degasses about 1 gigatonne (10^9 tonnes) of CO_2 (ref. 3), equivalent to 20 per cent of current anthropogenic output. The regression lines and intercepts of dissolved inorganic carbon, silicate and nitrate concentrations, measured in the upper 200 m of the EUZ, indicate that silicate availability regulates both carbon and nitrate uptake and fate. The slope of the highly significant regression between nitrate and silicate is 1, which coincides with the known requirement of these two elements by diatoms. The intercept represents the excess nitrate (about 4 mmol m^{-3}) that confers HNLC status on the EUZ.

Silicate is bound in mineral shells (biogenic silica) of no food value. In the EUZ model, the diatom shells are packaged in zooplankton faeces and exported wholesale from the surface layer. In contrast, nitrogen is selectively retained by the diatom grazers which, in the course of their metabolism, excrete ammonia; in turn the ammonia is assimilated by the photosynthesizing cyanobacteria ($<2 \mu\text{m}$) and small flagellates ($2\text{--}10 \mu\text{m}$) of the 'microbial loop'. This ubiquitous community (pico- and nanoplankton) is additionally composed of small protozoa that feed on the minute algae and bacteria, as well as on one another, which also release ammonia (see Fig. 2 of Dugdale and Wilkerson's paper, page 272). The ammonia-based production within this tightly geared community — termed regenerated production — contributed four-fifths of the total daily production (new plus regenerated) in the EUZ. Because of the low variability in nutrient and biomass concentrations observed over different cruises and seasons, the pelagic system of the EUZ seems to have settled into a steady state which can be likened to a silicate-limited chemostat.

The silicate-to-nitrate ratios in the other HNLC regions — the Southern Ocean and the subarctic Pacific — are much higher than in the EUZ, yet fairly similar community structures, total production rates and ratios of new-to-regenerated production are routinely measured there. Diatom growth is evidently not silicon-limited, so other factors, such as deep mixing, iron deficiency and heavy grazing pressure, have been invoked to explain the HNLC condition. An *in situ* iron fertilization experiment (IronEx II), carried out in the South Equatorial Current adjacent to the EUZ, yielded a spectacular diatom bloom⁴ and showed that iron availability limited new production in this region.

Dugdale and Wilkerson argue that, as the EUZ diatoms are already silicate-limited, adding iron should have no effect on them. They suggest that iron upwelling with the other nutrients in the EUZ is sufficient to meet the demands of the diatoms. In the iron-limited waters of the South Equatorial Current, iron-fertilized diatom growth would eventually be halted by silicate but

not nitrate exhaustion.

Yet the paradox persists. Virtually all phytoplankton species, including the picoplankton, are able to use nitrate; indeed, phytoplankton other than diatoms routinely exhaust nitrate in the surface waters of the non-HNLC ocean. So why does this not happen in the EUZ and other low-silicate HNLC regions? Differences in grazing pressure have been proposed as an explanation⁵. One widely held view is that the small algae of the microbial loop are kept in check by heavy grazing pressure, whereas diatoms, because of their larger size (and possibly also the protection offered by the silica shell), are less prone to being grazed by the smaller protozoa⁶. So relaxation of a limiting factor (such as iron) results in accumulation of diatom but not picoplankton cells. Whether continued iron fertilization will eventually lead to nitrate exhaustion by non-diatom phytoplankton in low-silicate HNLC regions remains to be tested.

In Dugdale and Wilkerson's steady-state EUZ model, the biomass-to-production ratio of the diatoms indicates that they were growing at least as fast as, if not faster than, the microbial algae. To maintain steady state, grazing pressure on the diatoms, presumably by copepods (zooplanktonic crustacea, equipped with mandibles edged with silica the better to crush diatom shells with), must have been similar to or even higher than that

on the microbial algae. The growth performance of the diatoms is all the more surprising as nitrate reduction requires energy and the mediating enzyme contains iron. Indeed, why diatoms can be so much more efficient than the other algae despite the nitrate handicap needs to be explained.

Balancing pelagic ecosystem budgets is still an art because we know so little about the abilities and predilections of the organisms and their interactions with one another⁵. Whatever the outcome of studies on the limiting factors in the various HNLC regions and their subsystems, the status of diatoms as key players will not be challenged. The workhorses running pelagic systems are recruited from this algal group: their new production not only fuels the food chains leading to fish but also provides the raw material driving the biological pump and ultimately the great biogeochemical cycles of the ocean. It is time we gained a better understanding of the properties that make diatoms so special. □

Victor Smetacek is at the Alfred-Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany.
e-mail: vsmetacek@awi-bremerhaven.de

1. Dugdale, R. C. & Wilkerson, F. P. *Nature* **391**, 270–273 (1998).
2. Doney, S. C. *Nature* **389**, 905–906 (1997).
3. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
4. Coale, K. H. *et al.* *Nature* **383**, 495–501 (1996).
5. Banse, K. *Oceanography* **7**, 13–20 (1994).
6. Banse, K. *ICES J. Mar. Sci.* **52**, 265–277 (1995).

Developmental biology

Hox gene variation and evolution

Axel Meyer

Both developmental and evolutionary biologists try to explain patterns in the diversity among organisms, and the Hox genes encode a class of transcription factors that have provided ample material for such discussions. Because they may be pivotal in specifying regional identity in body plans, differences in their expression could (at least partly) explain the evolution of animal phyla. The Hox genes are arranged in genomic clusters and, importantly, they are expressed in a spatially colinear fashion — anterior genes are expressed early in development and towards the front of the body, posterior genes later in development and in more distal portions of the body.

In invertebrates, only a single Hox gene cluster has been found (although it is split in *Drosophila*). The common ancestor of all chordates is surmised to have had a single cluster as well. This cluster is thought to have duplicated to four clusters (A–D) on different chromosomes, accompanying the increasing complexity of body plans during the evolution of vertebrates (Fig. 1). But a report by Prince *et al.*¹, shortly to appear in *Development*, is likely to cause some

questioning of this commonly held hypothesis — that genomic and morphological complexity are causally linked^{2–7}.

Using an experimental approach based on the polymerase chain reaction, Prince *et al.* unambiguously identified 34 Hox genes and determined their linkage with somatic-cell hybrids. Surprisingly, they found that the zebrafish has three Hox genes (*HoxC3*, *HoxA8* and *HoxB10*), with no direct mouse equivalents (Fig. 1). Moreover, the expression domains of the anterior Hox genes are partly overlapping, and restricted to a shorter anterior region. Possibly the most important finding is that the zebrafish has at least two additional Hox gene clusters for a total of six and not, as previously thought⁸, the typical vertebrate number of four. All of the genes on these additional clusters have probably not yet been discovered, and three are reported so far¹.

These two additional clusters lead us to question a simple, 'more clusters, more complexity' model of evolutionary diversification — in terms of phenotypic complexity, however measured, a zebrafish is probably not 50 per cent more complex than a mouse

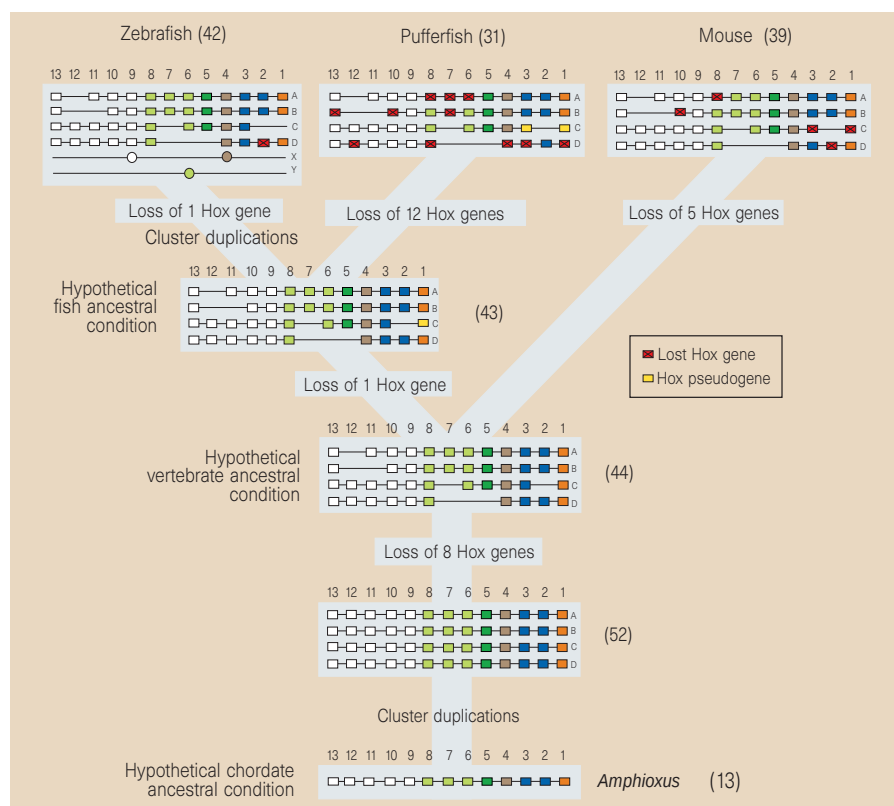


Figure 1 Evolutionary conservation of the Hox gene clusters in zebrafish, pufferfish, mouse and *Amphioxus*. The zebrafish Hox gene clusters are shown with the 34 genes confirmed by Prince *et al.*¹, and an additional eight, for the total of 42 that the zebrafish is likely to have. Ten gene losses differentiate the mouse from the pufferfish, but there are only three losses between the mouse and zebrafish, even though fish and tetrapods shared their last common ancestor about 400 million years ago. The common ancestor of human and mouse probably lived around 60 million years ago, but the architectures of their Hox gene clusters are identical. The losses in the *AbdA* (abdominal A)-related genes (light green) in the D-cluster probably occurred early in the history of vertebrates. (Figure drawn by Heike Haunstetter.)

or a human. The extra clusters cannot be explained by entire-genome duplications because, although polyploidy is known from other carp-like fish⁹ and is common in salmonids, zebrafish are diploid. It also seems unlikely that the additional clusters are remnants of a polyploid ancestral condition. Instead, the Hox-cluster duplications in zebrafish might be a unique evolutionary event. But such events may turn out to be common, at least in fish.

Prince and colleagues' work on zebrafish¹, combined with studies on the pufferfish¹⁰, now enables us to reconstruct the evolutionary history of the Hox gene clusters in vertebrates (Fig. 1). The initial chordate ancestral cluster of 13 Hox genes (the architecture that is still present in the cephalo-chordate *Amphioxus*⁵) probably duplicated in a three-step process, to form four complete clusters with a total of 52 genes. One phylogenetic study¹¹ indicates that the D-cluster is the most ancestral, and that the B- and C-clusters are the youngest. Hagfish and lamprey are phylogenetic intermediates between *Amphioxus* and more derived vertebrates such as zebrafish. So, if these fishes have only two or three clusters (which is not

precisely known), they would be more likely to contain a D-like Hox gene cluster than a B- or C-cluster. The suggestion that the D- and A-clusters are the oldest also seems to fit the observation that the D-cluster is the most 'deteriorated' of all (Fig. 1).

Following the principle of Dollo parsimony — which assumes that losses of genes are much more common and likely than independent evolutionary origins¹² — we can speculate which Hox genes might have been present in the common ancestors of vertebrates, tetrapods and fish (Fig. 1). Based on the genomic organizations available so far, the rates of evolution of Hox clusters do not seem to be constant. For example, whereas the zebrafish is likely to have lost only one Hox gene since it shared a common ancestor with the pufferfish (probably more than 200 million years ago), the losses along the pufferfish lineage were possibly several times faster (12 Hox genes were lost, if the zebrafish really has 42 Hox genes) (Fig. 1). Ignoring the additional clusters of the zebrafish, and estimating that it has 42 Hox genes, we find that 13 differences separate the zebrafish from the pufferfish. In the pufferfish, the *HoxC1* and *C3* genes are still



100 YEARS AGO

Messrs. Swan Sonnenschein and Co. announce that they will shortly publish a work, entitled "The Wonderful Century: its Successes and its Failures," by Dr. Alfred R. Wallace, F.R.S. The object of the volume is to give a short descriptive sketch of all the more important mechanical inventions and scientific discoveries which are distinctive of the nineteenth century. ... The author maintains that our century is altogether unique; that it differs from the eighteenth or seventeenth centuries, not merely as those differed from the centuries which immediately preceded them, but that it has initiated a new era, and that it may be more properly compared with the whole preceding historical period.

The January number of the *National Review* has an admirable article by Mr. Gerald Arbuthnot, entitled "In Defence of the Muzzle." The temperate spirit in which it is written, and the conscientious manner in which the statistics referred to have been collected, ought to materially strengthen the hands of those who are upholding the muzzling order for dogs, in the face of the selfish and short-sighted opposition which it is receiving from a certain section of the public. From *Nature* 13 January 1898.

50 YEARS AGO

A symposium arranged by the New York Academy of Sciences and held in December 1946 on "Nutrition in Relation to Cancer" covered a wide field and included a number of interesting articles which have now been published. ... Although it may seem disappointing that after so much study of cancer the fundamental cause or nature of it is unknown, the papers [in the symposium volume] show advances. The carcinogenic process can often be influenced by diet, which means that the process can be resolved into separate parts, and this must help in the understanding of the process. Many of the speakers at the symposium compared carcinogenesis to mutations. Both cancer and mutations can be induced in living organisms by similar agents. The hypothesis that cancer is a somatic mutation relates carcinogenesis to other biological changes and the stability of the nuclear and cytoplasmic genes. From *Nature* 17 January 1948.

recognizable, but they are only pseudogenes (genes that are not transcribed). They might have lost their function at different points in the evolution of fish, because *HoxC3*, at least, is still present in the zebrafish.

The apparent acceleration of genomic evolution along the pufferfish lineage might be correlated with accelerated morphological evolution. Pufferfish belong to one of the most morphologically derived groups of fish, and they lack ribs, pelvic fins and the pelvic girdle. Are the missing genes those that are no longer necessary because these structures have been lost during evolution? If so, the missing Hox genes in the pufferfish might also be absent in the other groups of fish which have secondary loss of pelvic fins (such as eels) or even tail fins (for example, the ocean sunfish *Mola mola*). Moreover, the loss of Hox genes might also be accompanied by the secondary loss (or simplification) of appendages in land vertebrates, such as in limbless amphibians, reptiles and whales.

What selective forces maintain or modify genomic organizations? The observation that Hox genes are clustered, and that the architecture of these clusters is highly conserved in evolution, has led to the suggestion that the regulatory elements that control expression of the Hox genes cannot be separated from these genes without jeopardizing their proper functioning and, possibly, determination of morphology along the antero-posterior axis. For vertebrates^{13,14} these ideas have been partially confirmed experimentally¹⁵, and this tight functional linkage might be particularly strong along the lineage that leads to reptiles and mammals (Fig. 1).

The long-standing question of whether

the evolution of genes or networks of interactions through regulatory elements drives most morphological diversification might, then, have different answers in different evolutionary lineages. In the most species-rich group of vertebrates — fish — organization of the Hox genes might not be completely constrained by interwoven regulatory networks, and differentiation might be driven by gene evolution. However, in the lineage that leads to reptiles and mammals, the driving force behind morphological diversification might have been newly evolving interactions in networks of regulatory elements¹⁶. □

Axel Meyer is in the Department of Biology, University of Konstanz, D-78457 Konstanz, Germany.

e-mail: axel.meyer@uni-konstanz.de

1. Prince, V. E., Jolly, J., Ekker, M. & Ho, R. *Development* **125**, 407–420 (1998).
2. Akam, M. *Phil. Trans. R. Soc.* **349**, 313–319 (1995).
3. Holland, P. W. H. & García-Fernández, J. *Dev. Biol.* **173**, 382–395 (1996).
4. Lewis, E. B. *Nature* **276**, 565–570 (1978).
5. Krumlauf, R. *Cell* **78**, 1991–202 (1994).
6. Duboule, D. & Morata, G. *Trends Genet.* **10**, 358–364 (1994).
7. García-Fernández, J. & Holland, P. W. H. *Nature* **370**, 563–566 (1994).
8. Misof, B. Y. & Wagner, G. P. *Mol. Phylog. Evol.* **5**, 309–322 (1995).
9. Buth, D. G., Dowling, T. E. & Gold, J. R. in *Cyprinid Fish, Systematics, Biology and Exploitation* (eds Winfield, I. J. & Nelson, J. S.) 83–126 (Chapman & Hall, London, 1991).
10. Aparicio, S. et al. *Nature Genet.* **16**, 79–83 (1997).
11. Bailey, W. J., Kim, J., Wagner, G. P. & Ruddle, F. H. *Mol. Biol. Evol.* **14**, 843–853 (1997).
12. Meyer, A. in *New Uses for New Phylogenies* (eds Harvey, P. H., Leigh Brown, A. J., Maynard Smith, J. & Nee, S.) 322–340 (Oxford Univ. Press, 1996).
13. Gérard, M. et al. *Genes Dev.* **10**, 2326–2334 (1996).
14. Gould, A. et al. *Genes Dev.* **11**, 900–913 (1997).
15. Mann, R. S. *BioEssays* **19**, 661–664 (1997).
16. Davidson, E. H., Peterson, K. J. & Cameron, R. A. *Science* **270**, 1319–1325 (1995).

stable isobars, in which more than one stable nucleus exists at a given mass number.

There are three basic nuclear processes that produce heavy elements: the s-, r- and p-processes, standing for slow, rapid and proton-rich. The abundances produced by each of these³ are shown in Fig. 1.

The s-process probably takes place in red giant stars that are burning helium, with a given nucleus capturing a neutron perhaps every few thousand years. But only fractions of a second separate successive photodisintegrations in the p-process or neutron captures in the r-process, implying that these processes must occur in supernova explosions. So some details of supernova explosions can be deduced from isotopic analysis of meteoritic material — in particular, interstellar grains that have undergone relatively little thermal processing during and since the formation of the Solar System.

One important recent discovery was that there are two different r-processes⁶: one responsible for the r-process nuclides up to about $A = 140$ (where A is the mass number), the other for nuclides above $A = 140$. The discovery stemmed from extinct radionuclides, which lived long enough to survive with measurable abundances from the time of their production to their injection into the forming Solar System, but not long enough to be measurably present now. They are detected through excesses of the daughter elements that result from their decay.

The abundance of ^{182}Hf in the early Solar System is consistent⁶ with the continuous production of the actinides in the Galaxy, with mixing to maintain a roughly constant abundance level on a timescale consistent with the mean life of ^{182}Hf — about 10^7 years. But trouble arises from two lighter radionuclides, ^{107}Pd and ^{129}I . If these were made by the same r-process that produced the

Isotope astrophysics

Two cradles for the heavy elements

A. G. W. Cameron

Where do the Solar System's heavy elements come from? We know that many of the elements heavier than iron must have been formed in supernovae; but exactly how? Taken together, three new papers (one on page 261 of this issue¹, and two to appear in the *Astrophysical Journal*^{2,3}) imply that two different types of supernova are responsible for elements in different mass regimes.

It is now some four decades since Suess and Urey⁴, largely working from analyses of meteorites, assembled a table of abundances of the elements that enabled nuclear physicists to identify the principal processes in stellar interiors that had produced those elements. The actual nuclear physics of those processes became reasonably well understood within the following few years. However, it has taken considerably longer for us to understand the astrophysical environ-

ments within stellar interiors in which many of those processes take place. The process whose environment has proved most elusive has been the r-process, in which neutron capture takes place on a very rapid timescale.

In the evolution of a star considerably more massive than the Sun, nuclear fusion reactions build toward products in which the binding energy per nucleon becomes maximized. This produces an abundance peak at ^{56}Fe . Making nuclides much heavier than this involves neutron capture, which generally produces nuclides on the neutron-rich side of the region of stable elements, known as the valley of beta stability. Only in the violence of a supernova explosion can the nuclides on the neutron-deficient side be produced, primarily by losing nucleons in photodisintegration. It is possible to distinguish these processes by examining the abundances of nuclei in the Solar System, particularly of

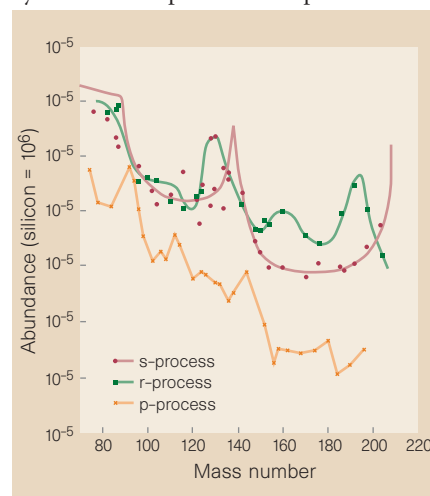


Figure 1 The solar abundances of the nuclides, as a function of mass number, showing the p-, s- and r-process contributions. For the s- and r-processes, average smooth lines have been drawn through the characteristic zig-zag patterns of the odd and even mass numbers. (From ref. 5.)

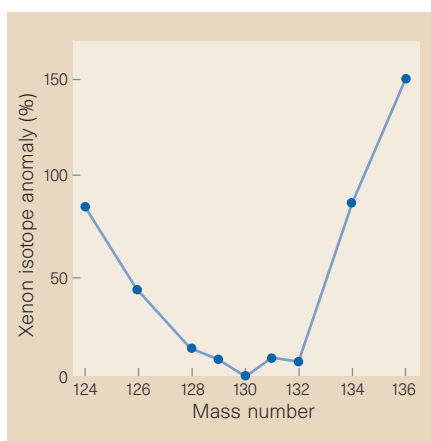


Figure 2 Overabundances of xenon from pre-solar diamonds, relative to solar values. (After Ott⁸.)

actinides, then they would be some two orders of magnitude more abundant than they are. So there must have been an interval of $\sim 10^8$ years between the last r-process production of ^{107}Pd and ^{129}I and their injection into the forming Solar System.

This is the basis for the claim that there are two distinct r-processes, with yields confined above and below about $A = 140$. The high-mass r-process must operate more frequently; a typical point in the interstellar medium receives products from it at roughly 10^7 -year intervals, but the low-mass-number r-process is much rarer, giving that point a contribution only every 10^8 years or so.

The newest results, which both confirm this idea and suggest how the two processes arise, come from interstellar diamond grains extracted from meteorites, and the anomalous isotopic abundances of their xenon and tellurium. These diamonds are tiny — only about one in a million contains a xenon atom, so the measurements must be made on bulk samples.

Figure 2 shows the overabundances in the principal xenon component contained in the diamonds (called XeHL). This pattern persists over a wide range of conditions of gas extraction from different samples at different temperatures⁷, indicating a common astrophysical source. Most striking are the high abundances of mass numbers 134 and 136 (r-process products), and 124 and 126 (p-process products). It has been suggested⁸ that this XeHL pattern is produced in a supernova environment by some kind of separation between those xenon isotopes with shorter-lived precursors and those with longer-lived precursors. But a new paper contradicts this idea.

Instead of xenon, Richter *et al.*¹ look at the tellurium in interstellar diamonds. It has large overabundances *only* at mass numbers 128 and 130, both r-process isotopes whose precursors live about an hour. The tiny overabundance ($< 4 \times 10^{-5}$) for the $A = 120$ p-process product, for example, is in striking

contrast to the xenon p-process products shown in Fig. 2. The time delay in forming $A = 120$ is also only about an hour, and the supernova p-process should⁹ have a yield at $A = 120$ comparable to those at $A = 124$ and 126 for xenon. So a separation based on precursor lifetimes can't explain the tellurium data. What can? To address this, we will have to look more closely at the extreme physical conditions occurring in supernovae.

The r-process is thought to involve neutrino-driven winds in the surface layers of a newly formed neutron star. A neutron star must be formed during the collapse process that produces a supernova, to supply the observed amount of energy. Enough energy is released so that the thermal energy, $E = kT$ (where k is Boltzmann's constant, and the temperature T is usually measured in MeV) is large compared with the rest-mass energy of the electron, and enough to fill the available phase space of quantum states with neutrinos. Inside the neutron star, the six flavours of neutrino — electron, muon and tau, and their corresponding antiparticles — interchange rapidly among themselves and with the radiation field.

The neutrinos leak out of the neutron star in seconds. Even when they are near the surface layers of the neutron star, however, the neutrinos continue to interchange with one another and with electrons, positrons and photons. This heats the surface matter, blow-

ing it off as a dense, neutron-rich wind. As it expands and cools, nuclear reactions take place that build up seed nuclei, and the r-process takes place on these seed nuclei as they cool further.

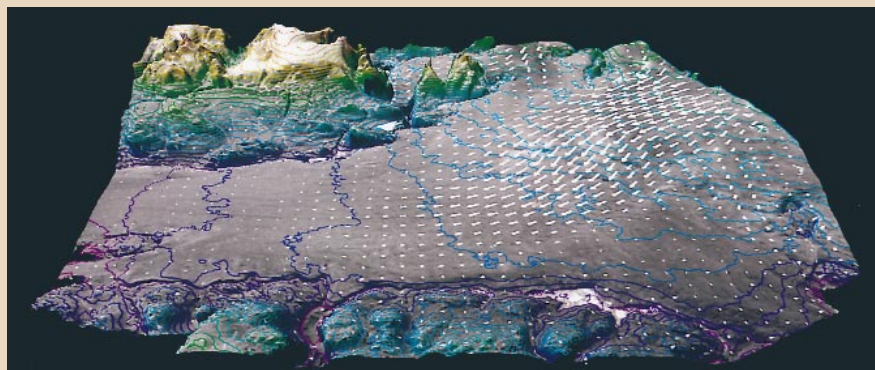
But this general description of the r-process can be applied to two very different cases involving high and low entropy — corresponding to high and low thermal energy.

In the hot, high-entropy case, the core of a massive star collapses to form a neutron star (and possibly further to form a black hole). For several years this has been the standard setting for the r-process, but simulations show that to match Solar System r-process abundances demands very high entropies indeed, corresponding to entropies of at least 400 (in units of Boltzmann's constant per baryon). Theory balks at this: it appears to be impossible^{10,11} to reach an entropy as high as 200. Perhaps the theorists should concentrate less on increasing the numbers of neutrons and look at reducing the numbers of seed nuclei, so that the existing neutrons can carry the r-process build-up to higher mass number.

Nevertheless, neutron stars with massive stellar precursors seem essential to understand both the r-process and the p-process that produced the XeHL in Fig. 2. The p-process is thought to operate mainly in the oxygen–neon envelope of such a massive star⁹. Can the high-entropy picture then be

Satellite mapping

3-D glacial flow



Most satellite maps are only two-dimensional. But in the past few years, synthetic-aperture-radar (SAR) images from artificial satellites have been combined into interferograms, to produce high-resolution topographic maps of the Earth's surface and spectacular maps of surface deformation over active faults, volcanoes and glaciers. On page 273 of this issue, Johan Mohr and colleagues present a further improvement of this technique (Mohr, J. J., Reeh, N. & Madsen, S. N. *Nature* 391, 273–276; 1998).

Instead of mapping only the radar line-of-sight component of ice flow, as in

conventional SAR interferometry, they have taken advantage of the differing angles afforded by ascending and descending satellite orbits to map the full three-dimensional flow pattern of Storstrømmen, a large outlet glacier in northeastern Greenland. Above is a perspective view of this flow field looking from the east, with white arrows representing the glacial flow. The slowing near the end of the glacier (left of figure) is unusual, but expected for a glacier that has recently surged. This method can be used to assess the volume changes of land ice masses.

John VanDecar

adapted to provide two distinct r-processes? One suggestion² is that the formation of a neutron star as an end point might generate enough neutrons to drive the r-process up into the actinide region, whereas if the end-point is instead a black hole, then the neutrons would be prematurely cut off and only the low-mass-number range of products would be produced. Although this is an attractive idea, it doesn't predict the tellurium results, in which there is no trace of a p-process yield. As the supernova shock would be launched before the black hole ended neutron production, the p-process products should be there.

But a low-entropy r-process has now been discovered³ that could give products primarily in the lower range of mass numbers. Many stars of 6–12 solar masses exist in binary systems. At some point, these stars expand, transfer most of their hydrogen to their companions, and become white dwarfs. Eventually they accrete material back from their companions, and some of them grow in mass until the Chandrasekhar limit is reached, at which point the star collapses to form a neutron star. This is accretion-induced collapse (AIC), and results in so-called type-1a supernovae.

The entropy per particle in AIC is only about 3–5. As an entropy of at least 8 would be needed to completely break up the nuclei into nucleons, the star must retain very large numbers of seed nuclei. So the ratio of neutrons to seed nuclei would be relatively small, and this would favour a final distribution of r-process products in the lower range of mass numbers.

Further, the lack of a massive envelope means that AIC events do not lead to a p-process, and so could fit the tellurium results.

How were these isotopes implanted in the interstellar diamonds? Industrial diamonds are often produced by chemical vapour deposition in a cooling flow; and presumably something similar happens when a supernova shock wave encounters a carbonaceous layer. In a massive supernova, such a layer would be in the outer envelope of the star, whereas in an AIC supernova the carbon is likely to be in the material from the companion star flowing onto the white dwarf in the binary pair. But this requires further exploration.

Nevertheless, a consistent picture has emerged in which rare, low-entropy AIC supernovae inject the lighter r-process isotopes, and more common core-collapse supernovae take over for isotopes heavier than about $A = 134$. □

A. G. W. Cameron is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

1. Richter, S., Ott, U. & Begemann, F. *Nature* **391**, 261–263 (1998).
2. Qian, Y.-Z., Vogel, P. & Wasserburg, G. J. *Astrophys. J.* (in the press).
3. Baron, E., Cowan, J. J., Rogers, T. & Gutierrez, K. *Astrophys. J.* (submitted).
4. Suess, H. E. & Urey, H. C. *Rev. Mod. Phys.* **28**, 53–74 (1956).
5. Cameron, A. G. W. in *Essays in Nuclear Astrophysics* (eds Barnes, C. A., Clayton, D. D. & Schramm, D. N.) 23–43 (Cambridge Univ. Press, 1982).
6. Wasserburg, G. J., Busso, M. & Gallino, R. *Astrophys. J.* **466**, L109–L113 (1996).
7. Huss, G. R. & Lewis, R. S. *Meteoritics* **29**, 791–810 (1994).
8. Ott, U. *Astrophys. J.* **463**, 344–348 (1996).
9. Rayet, M. *et al.* *Astron. Astrophys.* **298**, 517–527 (1995).
10. Qian, Y.-Z. & Woosley, S. E. *Astrophys. J.* **471**, 331–351 (1996).
11. Hoffman, R. D., Woosley, S. E. & Qian, Y.-Z. *Astrophys. J.* **482**, 951–962 (1997).

Structural biology

A mechanism for all polymerases

Thomas A. Steitz

Possibly the earliest enzymatic activity to appear in evolution was that of the polynucleotide polymerases — the ability to replicate the genome accurately being a prerequisite for evolution itself. Thus, one might anticipate that the mechanism by which polymerases work would be both simple and universal. Further, these enzymatic scribes must faithfully copy the sequences of the genome into daughter nucleic acid, or the information contained within will be lost for ever. Finally, replicative DNA polymerases are highly processive, traversing the whole genome of a virus DNA without falling off.

On pages 251 and 304 of this issue, Doublé *et al.*¹ and Kiefer *et al.*² provide significant insights into the catalytic mechanism, fidelity and processivity of DNA polymerases. In a detailed crystal structure at 2.2 Å resolution, Doublé *et al.* have captured the DNA polymerase that replicates bacteriophage T7 DNA, along with the *Escherichia coli* thioredoxin in the act of adding a

deoxynucleoside triphosphate (dNTP) to a primer–template DNA. The incoming dNTP (actually dideoxy here) is accompanied by two magnesium ions, which are bound to the phosphates of the nucleotide and to two aspartic-acid residues that are widely conserved among DNA and RNA polymerases. To prevent the reaction from occurring in the crystal, the authors used a dideoxynucleotide at the primer terminus whose 3' OH, when present, would interact with one of the two metal ions.

This structure supports a 'two-metal-ion' mechanism of nucleotide addition, which was proposed³ by analogy to the nearly identical mechanism of the 3'–5' exonuclease of DNA polymerase I (Fig. 1). In this mechanism, metal ion A lowers the affinity of the 3' OH for the hydrogen, facilitating the 3' O[−] attack on the α-phosphate. Metal ion B assists the leaving of the pyrophosphate, and both metal ions stabilize the structure and charge of the expected pentacoordinate transition state. This two-metal-ion-catalysed

mechanism could act in an RNA or DNA polymerase made from RNA rather than protein and, thus, could function in an all-RNA world.

An identical two-metal-ion polymerase mechanism seems to be used by another, non-homologous DNA polymerase — mammalian DNA polymerase β (Pol β) — presumably as a consequence of convergent evolution. Kraut and co-workers⁴ found a similar arrangement of primer–template DNA, ddNTP and magnesium ions in Pol β. Indeed, when Doublé *et al.*¹ superimposed the primer–template DNAs as bound to these two polymerases, they found that the nucleotide, the two magnesium ions and their two aspartic-acid ligands superimpose likewise, in spite of the differences between the two protein structures. The conservation of catalytic-domain structure and liganding of the metal ions by two aspartic-acid residues, in structures of the B (Pol α)-family DNA polymerases⁵, T7 RNA polymerase⁶, human immunodeficiency virus reverse transcriptase⁷ and poliovirus RNA-dependent RNA polymerase⁸, implies that all families of polynucleotide polymerases may use the same, two-metal-ion mechanism.

One source of the impressive fidelity with which polymerases copy DNA is uncovered in the co-crystal structure of *Bacillus stearothermophilus* DNA polymerase large fragment with primer template, solved by Kiefer *et al.*², and that of the T7 polymerase complex¹. Both provide details of a sequence-independent molecular recognition of correctly formed base pairs. Protein side-chains interact, in the minor groove,

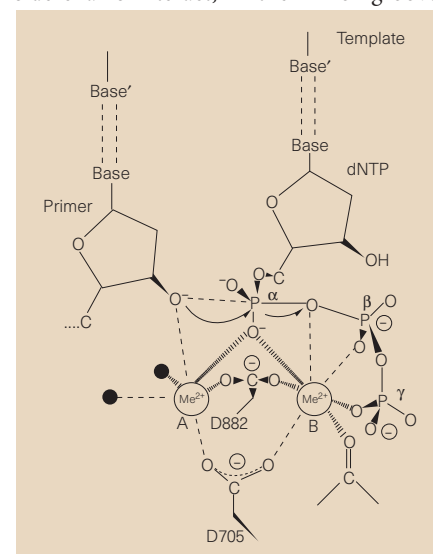


Figure 1 The two-metal-ion mechanism of polynucleotide polymerases³ in the context of the T7 DNA polymerase–substrate complex solved by Doublé *et al.*¹. Two divalent metal ions, A and B, are ligated to enzymes of the *Escherichia coli* DNA polymerase I family by aspartic-acid residues 705 and 882. The black circles are water molecules bound to metal ion A. (Modified from ref. 12.)

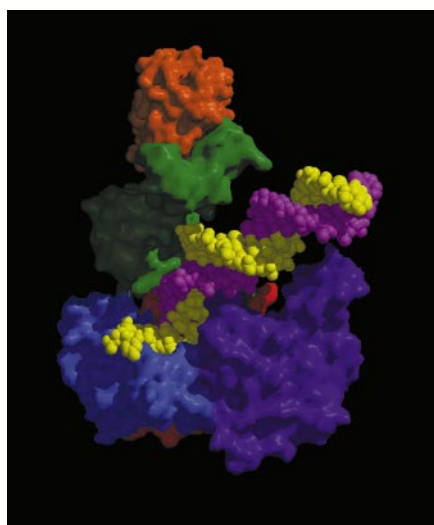


Figure 2 Structure of the T7 DNA polymerase, solved by Doublé *et al.*¹, locked in synthesis mode. The fingers (blue), thumb (green) and palm (red) domains are shown, as well as the exonuclease (purple) and *Escherichia coli* thioredoxin (orange).

with the hydrogen-bond acceptors that lie in identical positions in all four Watson–Crick base pairs. These interactions can occur only with correct Watson–Crick base pairs, and not with mismatches. The DNA near the primer terminus has the widened minor groove of an A-form structure (as observed with other polymerase–DNA complexes^{4,9,10}), and this facilitates access to the minor groove. The base of the incoming dNTP and its template partner fit snugly into a pocket that accommodates only Watson–Crick pairing¹.

The crystal structure of the T7 polymerase complex includes the *E. coli* thioredoxin, which renders the polymerase so processive that it can replicate the entire T7 genome before falling off¹¹. The thioredoxin subunit interacts with a polymerase domain termed the ‘thumb’ (Fig. 2). It is adjacent to the duplex DNA product in a region 12–18 base pairs from the primer terminus, where the DNA in this crystal becomes disordered. Consequently, the mechanism by which thioredoxin confers processivity remains uncertain. The thioredoxin, which is flexibly tethered to a binding loop in the thumb, could swing across the DNA-binding groove to encircle the DNA¹. Or, less likely, it could simply extend the DNA-binding site¹. Kiefer *et al.*² observed that their crystals of *Bacillus* fragment are catalytically active, and can undergo several rounds of nucleotide incorporation. Thus, the processive sliding that is thought to accompany the translocation step in DNA synthesis can occur within the confines of the crystal.

Comparison of the Pol β or T7 DNA polymerase ternary-complex structures with the corresponding binary complexes or apo-enzyme structures shows that binding

of both substrates results in a dramatic change in the orientation of the ‘fingers’ domain. In the T7 enzyme¹, the whole domain rotates about 40° (relative to the primer template), providing the proper binding site for the incoming dNTP and its template partner. In the *Bacillus* fragment binary complex², the template base is not positioned to pair with the incoming dNTP. The relationship of this conformational switch to fidelity and translocation is worth further examination.

These two co-crystal structures^{1,2}, as well as those of *Thermus aquaticus* DNA polymerase⁹ and HIV reverse transcriptase¹², bind primer–template DNA to their catalytic (‘palm’) domains in a similar fashion. But this is not the way that was proposed by Pelletier *et al.*⁴ from their co-crystal structure of Pol β . Because Pol β is not homologous to these enzymes, it can be compared only by alignment of the bound primer–template DNAs. However, striking analogies can then be seen within the non-homologous structures of the ‘fingers’ and ‘thumb’ domains, which bind the dNTP and primer–template. By comparing the functional complexes of

these structurally diverse polymerases we should, therefore, be able to identify those structural elements that are essential for the polymerase reaction. □

Thomas A. Steitz is in the Department of Molecular Biophysics and Biochemistry, and the Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA.

1. Doublé, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. *Nature* **391**, 251–258 (1998).
2. Kiefer, J. R., Mao, C., Braman, J. C. & Beece, L. S. *Nature* **391**, 304–307 (1998).
3. Steitz, T. A. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).
4. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. *Science* **264**, 1891–1903 (1994).
5. Wang, J. *et al. Cell* **89**, 1087–1099 (1997).
6. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B. C. *Nature* **364**, 593–599 (1993).
7. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. *Science* **256**, 1873–1790 (1992).
8. Hansen, J. L., Long, A. M. & Schultz, S. C. *Structure* **5**, 1109–1122 (1997).
9. Eom, S. H., Wang, J. & Steitz, T. A. *Nature* **382**, 278–280 (1996).
10. Jacobo-Molina, A. *et al. Proc. Natl. Acad. Sci. USA* **90**, 6320–6324 (1993).
11. Tabor, S., Huber, H. E. & Richardson, C. C. *J. Biol. Chem.* **262**, 16212–16223 (1987).
12. Brautigam, C. & Steitz, T. A. *Curr. Opin. Struct. Biol.* (in the press).

Smart materials

Off and on reflection

Philip Ball

The architect’s palette has been enriched by smart materials, which enable a building to adjust its environment to the prevailing conditions. At a meeting last month*, Peter Duine (Philips Research Labs, Eindhoven) described a tantalizing addition to the palette: windows that can be reversibly switched from a transparent to a mirror state.

The appeal of a switchable mirror might be obvious to gimmick addicts, but are there more practical attractions in such devices? For privacy purposes it beats drawing the curtains, but one can also imagine applications in energy-conscious architectural design — using switchable mirrors as elements of smart systems that control the amount of sunlight admitted through a window. And as a means to influence the path of light beams, the mirrors might be put to good use in display devices and interior lighting systems, particularly when the light sources are hard to reach and adjust.

The new work, which brings these technological goals closer to fruition, stems from a report in 1996 of switchable optical properties in yttrium and lanthanum¹. The underlying principle is a transition between a reflective metallic state and a translucent insulating state, induced by converting a thin film of a metal to its ionic hydride. All of the alkali, alkaline-earth and rare-earth met-

als form hydrides on exposure to hydrogen gas, whereupon ionization of the metal can empty the conduction band and make the compound insulating. A thin enough (sub-micrometre) film may then be transparent.

For trivalent rare-earth metals like yttrium and lanthanum, however, the situation is slightly more complicated. In particular, the metals can exist in both a divalent and a trivalent state, corresponding to the hydrides MH₂ and MH₃. Three different phases occur on exposure to hydrogen gas: the α -phase, a solution of a small amount of interstitial hydrogen in the metal; the β -phase (the dihydride); and the γ -phase (the trihydride). The two hydride phases (β and γ) can easily be interconverted by altering the hydrogen pressure, as the gas can readily diffuse in and out of the thin film. The dihydride retains a partially filled conduction band, and so provides the metallic mirror, whereas the trihydride is transparent.

To make a device, the thin metal films deposited on glass are enclosed in a sealed cell after being coating with a 20-nm film of palladium to prevent oxidation. Palladium is a renowned hydrogen ‘sponge’, and so hydrogen gas introduced into the cell has easy access to the rare-earth metal. Graded-diffusion experiments, in which the gas is permitted access to the films from one end only, show the formation of the γ -, β - and α -phases, successively further away from the hydrogen source region.

* Fall Meeting of the Materials Research Society, Boston, 1–5 December 1997.



Figure 1 Transparent, dark and reflective states of a $Gd_{30}Mg_{70}$ alloy. The states can be switched by pumping hydrogen in and out of a cell that contains the alloy film.

The early work¹ on yttrium demonstrated the principle but left much to be desired in terms of optical properties. For one thing, yttrium dihydride is not fully reflective, but has a spectral transmission window: it is dark red in transmission, blue in reflection. The trihydride, meanwhile, is not perfectly transparent, but has a yellowish tinge, because its bandgap is not large enough to prevent absorption of high-energy blue photons.

In both respects, a magnesium-based system would do better. But the transition between metallic magnesium and the insulating dihydride (there is no trihydride in this case) is not readily reversible, and the diffusivity of hydrogen in magnesium is low (so the formation of the hydride is slow). But the Philips team have found a compromise, in which the good optical properties of magnesium are combined with the good switchability of rare-earth elements. Their favoured material at present is an alloy of magnesium and gadolinium²: this alloy can be switched in about one second, has no transmission window in the metallic state, and for magnesium contents of around 50% the insulating state is virtually colourless. The transparent phase forms at high (>1 atmosphere) hydrogen pressures, and the metallic reflective state at low pressures (for which the

metal/hydrogen ratio in the film is around 0.8). In between these two extremes a third phase is found which is dark and has low reflectivity (Fig. 1).

But devices that pump hydrogen gas are not destined to find a thriving market. To make a user-friendly package, the hydrogen will need to be transported in a condensed medium. Duine also reported a 'wet' electrochemical cell, in which the hydrogen is shuttled back and forth between the gadolinium–magnesium alloy film at one electrode, and a liquid electrolyte of potassium hydroxide in contact with another electrode, through the reduction of water (ref. 3): $H_2O + e^- \rightarrow H + OH^-$. What is really needed, however, is a solid-state cell in which the hydrogen is kept in some 'storage' layer and is transported through a solid electrolyte. 'Solid' proton conductors are widely known already, such as the polymer Nafion that is used in fuel cells. But a solid electrolyte that can deliver hydrogen for uptake as anions may be harder to find. Until then, we will have to keep drawing the curtains. □

Philip Ball is an associate editor of *Nature*.

1. Huijberts, J. N. *et al.* *Nature* **380**, 231–234 (1996).
2. Van der Sluis, P., Ouwerkerk, M. & Duine, P. *Appl. Phys. Lett.* **70**, 3356–3358 (1997).
3. Notten, P., Kremers, M. & Griessen, R. *J. Electrochem. Soc.* **143**, 3348–3353 (1996).

Tumour suppressors

Teaming up to restrain cancer

Moshe Oren

Foremost among the protective mechanisms that have evolved to reduce the likelihood of cancer are the tumour-suppressor genes. As long as they are fully active, these genes serve as potent buffers against tumour progression. Only when one (or more) of them becomes defunct is the gate opened for uncontrolled cell multiplication and cancer.

Perhaps the best known of all tumour-suppressor genes is *p53* (ref. 1), which is altered in about half of all human tumours, making it the most frequent target for genetic alterations in cancer. Presumably, loss of the normal *p53* protein facilitates the emergence of malignant traits. But, as Garkavtsev

*et al.*² report on page 295 of this issue, *p53* does not restrain cancer single-handedly. Rather, it seems to require an intimate partner — a protein encoded by another tumour-suppressor gene, *ING1*.

The *ING1* gene was identified through a cloning strategy aimed at genes whose expression is selectively reduced in cancer cells³. Overexpression of *ING1* was found to block cell proliferation³ and to increase programmed cell death in the absence of survival factors⁴. When expression of *ING1* is quenched using antisense RNA, cellular replicative senescence is delayed, whereas unrestrained proliferation and illegitimate cell survival are promoted. All of these prop-

erties are consistent with *ING1* behaving like a tumour suppressor, underscoring a striking functional similarity between *ING1* and *p53*.

Prompted by these parallels, Garkavtsev *et al.* took a closer look at the relationship between the two proteins, *p33^{ING1}* and *p53*. They now report that this relationship extends a long way — in fact, each of the two proteins seems to require the other in order to exert its inhibitory effects. More specifically, *p33^{ING1}* cannot block the proliferation of cells lacking *p53*, nor can it render *p53*-deficient cells sensitive to chemotherapy. Conversely, if production of *p33^{ING1}* is suppressed using antisense RNA, cells can escape *p53*-mediated growth inhibition. So *p53* and *p33^{ING1}* seem to operate as a close team, presumably specializing in tumour suppression.

How is the team effort coordinated? The data indicate that *p33^{ING1}* and *p53* physically interact to form a specific protein–protein complex (see Fig. 1, overleaf). *p53* is mainly a sequence-specific transcription factor and, due to its interaction with *p33^{ING1}*, it becomes a more effective transcriptional activator. For example, *p33^{ING1}* can increase *p53*-dependent activation of the *p21/WAF1* promoter. The *p21^{WAF1}* protein potentially blocks the cell-cycle machinery and mediates *p53*-dependent growth arrest¹. Elevated production of *p21^{WAF1}* could, therefore, largely account for the ability of *p33^{ING1}* to block cell proliferation in concert with activated *p53*.

So far, the cooperation between *p53* and *p33^{ING1}* has been shown for only one gene (*p21/WAF1*) and, primarily, for one biological outcome — growth arrest. The pathway that leads to *p53*-mediated apoptosis is distinct from that leading to growth arrest, and probably involves the activation of a different subset of target genes. Some of these genes may promote apoptosis through the generation of reactive oxygen species⁵, and it remains to be seen whether *p33^{ING1}* cooperates with *p53* here too. But the answer will probably be positive: *p33^{ING1}* also exerts pro-apoptotic effects⁴, and it cooperates with *p53* to confer sensitivity to chemotherapy², a treatment that works mainly by inducing apoptosis of cancer cells.

In the experimental systems used by Garkavtsev *et al.*², the presence of *p33^{ING1}* seems to be obligatory for the inhibitory effects of *p53*. However, these systems are rather artificial and prone to exaggeration. So, within the human body, at least some of the physiological tumour-suppressor activities of *p53* may occur independently of *p33^{ING1}*. Nevertheless, if the loss of *p33^{ING1}* compromises the function of *p53* only slightly, this may provide emerging cancer cells with a selective growth advantage. And even a small advantage might make the difference between cancer and health.

As is often the case with crucial bioregula-

tory processes, the story is probably far more complicated. For example, at least some anti-proliferative activities of p53 rely on another candidate tumour-suppressor gene, *IRF1* (ref. 6). Cooperative activation of the *p21/WAF1* gene is implicated in this case, too. However, the mechanism may be distinct from that described for p33^{ING1}, and there may not be direct protein-protein association between p53 and IRF1. Furthermore, p53 can physically associate with yet another tumour-suppressor gene product — the WT1 protein⁷. Perhaps all associates belong to the same team or, maybe, p53 chooses different partners under different circumstances.

What are the practical implications of the new findings? For one, the concentration of p33^{ING1} might determine the extent to which a given tissue can mount an effective p53 response on exposure to stress. This could explain the puzzling observation that, unlike the promiscuous activation of p53 in cultured cells, only a limited number of tissues within a normal organism show strong p53 activation in response to DNA damage. An even smaller number show the expected biological effects⁸.

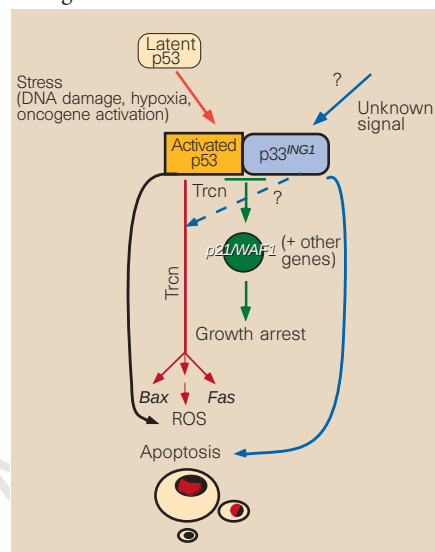


Figure 1 p33^{ING1} and p53 as mediators of cell fate. p53 can arrest cell proliferation by activating a set of target genes, among which *p21/WAF1* is pivotal. p53 can also induce cell death through apoptosis. This outcome relies, at least in part, on the transcriptional activation (Trcn) of death-promoting genes such as *Bax* and *Fas/Apo1*, and of genes involved in the production of reactive oxygen species (ROS). To exert its cellular effects p53 must first be biochemically activated, typically entailing an increase in cellular levels of p53. As Garkavtsev *et al.*² now report, many of p53's activities are carried out in cooperation with p33^{ING1}, presumably through a physical interaction between the two proteins. But it is not known which signals modulate the activity of p33^{ING1}, and whether p53-mediated apoptosis also requires p33^{ING1}.

The findings of Garkavtsev *et al.*² may also relate to the important question of whether p53 is fully functional in the 50 per cent of human tumours that retain a wild-type p53 gene. Certain types of tumour inactivate their p53 through increased expression of the p53-binding protein Mdm2 (ref. 9). So loss of p33^{ING1} function is another potential mechanism for inactivation of p53 in cancer cells. Moreover, in a limited analysis of three neuroblastoma cell lines, one line harboured a rearranged *ING1* gene resulting in a truncated protein³. Many more lines must be analysed before firm conclusions can be drawn, but it is remarkable that neuroblastoma cells often contain high levels of wild-type p53 (ref. 10) — a counter-intuitive behaviour given the highly malignant nature of these cells.

Despite being highly abundant, the p53 in neuroblastoma cells is functionally defective¹¹ (although this is somewhat controversial). The possibility that these cells carry a defect in p33^{ING1} might explain their unusual immunity to p53. Furthermore, the excess p53 accumulates in the cytoplasm of neuroblastoma cells, rather than in the nucleus, raising the possibility that p33^{ING1} might enable efficient nuclear translocation of p53.

A growing effort is now being invested in developing p53-based cancer gene therapies¹². If p33^{ING1} is essential for the effective anti-tumour action of p53, we may need to determine the status of p33^{ING1} in every tumour before subjecting the patient to such experimental therapy. There is already reason to suspect that *ING1* may be deleted in some head and neck tumours, which are prime targets for clinical trials in the near future¹². What is badly lacking now is an extensive assessment of alterations in p33^{ING1} — at the level of the gene as well as the protein — in human cancer. If such alterations are encountered at a significant frequency, we can take it for granted that p33^{ING1} will soon follow in the footsteps of its better-known team-mate and, like p53, will become a hot item in biomedical research. □

Moshe Oren is in the Department of Molecular Cell Biology, The Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: lioren@dapsas1.weizmann.ac.il

- Levine, A. J. *Cell* **88**, 323–331 (1997).
- Garkavtsev, I. *et al.* *Nature* **391**, 295–298 (1998).
- Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. *Nature Genet.* **14**, 415–420 (1996).
- Helbing, C. C., Veillette, C., Riabowol, K., Johnston, R. N. & Garkavtsev, I. *Cancer Res.* **57**, 1255–1258 (1997).
- Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. *Nature* **389**, 300–305 (1997).
- Tanaka, N. *et al.* *Nature* **382**, 816–818 (1996).
- Maheswaran, S., Englert, C., Bennett, P., Heinrich, G. & Haber, D. A. *Genes Dev.* **9**, 2143–2156 (1995).
- MacCallum, D. E. *et al.* *Oncogene* **12**, 2575–2587 (1996).
- Momand, J. & Zambetti, G. J. *Cell. Biochem.* **64**, 343–352 (1997).
- Davidoff, A. M., Pence, J. C., Shorter, N. A., Iglehart, J. D. & Marks, J. R. *Oncogene* **7**, 127–133 (1992).
- Moll, U. M. *et al.* *Mol. Cell. Biol.* **16**, 1126–1137 (1996).
- Barinaga, M. *Science* **278**, 1036–1039 (1997).

Daedalus

Bacterial diplomacy

No man is an island, said John Donne. Each of us is a community of billions, nearly all of them bacteria. The human skin, gut, mouth, vagina, etc., all harbour complex ecologies of many different micro-organisms, all living in seeming harmony. But how is this harmony maintained between selfish species? Moulds such as *Penicillium* take selfishness to extremes, and attack their bacterial rivals with deadly antibiotics. But more subtle discouragements must be more widespread. Thus, while it is quite easy to contract either syphilis or gonorrhoea, it is much harder to get both. Each disease seems to protect against the other. Similarly, few people carry more than one tapeworm; the first arrival somehow 'warns off' later ones. And invading bacteria are also usually warned off by our permanent flora. Clearly, says Daedalus, the chemical signalling behind our bacterial harmony is a set of gentlemen's agreements rather than gang warfare.

So DREADCO bacteriologists are now studying these agreements. They are culturing simple mixtures of gut and skin organisms, observing their equilibria and isolating the metabolites they produce. A signalling metabolite would be expected to shift the equilibrium of a culture to which it was added in excess, or from which it was rigorously excluded. Gradually the project should build up a library of such signalling metabolites, and discover their scope and mode of action.

The final goal is a new and milder form of antibacterial therapy. Many diseases, from vague inflammations and stomach upsets to sinusitis and genital infections, are upsets of bacterial equilibria. The medics move in with antibiotics, kill off goodies and baddies alike, and leave the field open for a new colonization. With luck it re-establishes the original equilibrium. But how much neater to drop in a nicely chosen mixture of chemical signals, warning the rioters to calm down, telling the persecuted to stand up for their rights, intimidating the foreigners, and thus restoring the equilibrium directly!

Yet Daedalus may have been anticipated. He now wonders if our immune system keeps our benevolent microbial ecologies in balance by multiple chemical signals. If so, all multicellular species may use such signals in their negotiations with bacteria. Hence, possibly, the bafflingly complex multi-component nostrums of the old herbalists and animal-product therapists.

David Jones

Goldsworthy's genera

Andy Goldsworthy sets out to explore some of the recurring forms in the world around us, using a variety of media drawn from nature itself – snow and sand, twigs and thorns.

Martin Kemp

Round holes, often of a tangible blackness, are fringed by miraculous aureoles of glowing autumn leaves. Sometimes the dark apertures are bounded by pebbles, graded in size and colour, while at other times they are hedged in by nests of bleached sticks. Long, thin fissures run jaggedly across layered carpets of leaves and through apparently random arrays of stones. Craters and snaking ridges of sand and soil speak of an organic geometry that blends the rectilinear and curvilinear without dissonance.

Low, vertiginous arches are engineered from raw plates of slate or irregular paving stones of ice, each discrete unit, courtesy of gravity, mutually squeezing its companions into stable unity in the absence of adhesive. A sharp star of shining silvery icicles tiptoes on needle-thin points across abrasive rocks. Cut slits transform a wall of frozen snow into an arctic Stonehenge.

Oval stones and angular rocks balance miraculously over single points. Twigs and thin sticks are woven into spiral bowers or giant cobwebs, as unexpected in pattern as they are paradoxical in their stability. A curved, tapering cornucopia, freshly green, is composed from a continuous spiral of sweet chestnut leaves, its seams stitched with thorns.

The British 'sculptor' Andy Goldsworthy is unrivalled in range and skill as a 'nature artist'. He works as a collaborator with natural phenomena during the passing seasons, feeling deep into the essence of soil, stone, water, ice, flower, leaf, stem, stick and trunk.

He is a maker in the manner of the bowerbird, spider, swallow, bee, shellfish — or, in human terms, the Eskimo, dry-stone waller or oriental master of Zen gardens. Tensile strength is conjured from fragile delicacy, and stable unity is compounded from inchoate members.

In the illustrations, Japanese maple leaves, startlingly carrot in hue, metamorphose from a floating chain one day to become the next a fringed hole supported by a woven briar ring. And blades of grass, creased and arched into a snake of expanding girth, are pinioned to the soil by ranks of thorn rivets.

Goldsworthy creates genera of forms — spheres, rings, holes, arches, spires, snakes and fissures — which cut across (sometimes literally) the conventional taxonomies of genus and species. Through sight, touch and motion he searches out, like a latter-day



Goldsworthy's *Japanese Maple Leaves, Supported by a Woven Briar Ring* (above), 1987 (Ouchiyama, Japan).



Blades of Grass, Creased and Arched, Secured with Thorns (left), 1988 (Penpont, Dumfriesshire).

D'Arcy Wentworth Thompson, morphological commonalities which arise from the material and processes of nature. For example, Goldsworthy writes: "Some works have the qualities of snaking but are not snakes... The snake has evolved through a need to move close to the ground, sometimes below and sometimes above, an expression of the space it occupies. This is a potent recurring form in nature which I have explored through working in bracken, snow, sand,

leaves, grass, trees, earth. It is the ridge of a mountain, the root of a tree, a river finding its way down a valley."

Across diversity, Goldsworthy sees commonalities, and in complexities he intuitively shared shaping. An instinctive dialogue with the sciences of order and disorder is being established. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Mudskippers store air in their burrows

Mudskipper fishes can maintain their metabolism while they are confined in mudflat burrows filled with oxygen-depleted water, and their eggs, deposited in the burrows, can develop under severely hypoxic conditions¹. How they cope with such conditions has been unclear. We report here that a mudskipper species *Periophthalmodon schlosseri* (Fig. 1) accumulates air in its burrows. This behaviour seems to be an adaptation to provide oxygen for burrow-dwelling fish and for embryos developing in the burrows.

Mudskippers are exceptional among fishes in their amphibious behaviour. Although many details are known about mudskipper terrestrial behaviour^{2,5} we have little information about their subsurface life in mudflat burrows other than that these are used for refuge and spawning²⁻⁴. *P. schlosseri*, one of the largest mudskippers in the world, inhabits intertidal mudflats in Southeast Asia^{6,7}. Our field studies in Malaysia have shown that the *P. schlosseri* burrows occur in the high intertidal zone and consist of a largely vertical shaft about 8 cm in diameter which connects to one or more horizontal tunnels (Fig. 2a). The maximum burrow depth recorded was 125 cm. The burrows were always filled with water and the water level was unaffected by tidal oscillation. The dissolved oxygen concentration of burrow water varied from almost none to more than 80% of that of air-equilibrated water at the surface. It declined sharply with depth, becoming less than 3% below 50 cm. Severely hypoxic conditions in mudflat burrows have been documented for other mudskipper species^{8,9}.

When we walked around the area where *P. schlosseri* burrows were found, we discovered that compression of the horizontal burrow tunnel expelled gas through the vertical shaft. We measured the volume and respiratory contents of the gas from 18 burrows to determine its possible relationship to mudskipper respiration. From 14 burrows we measured a gas volume of 55–630 ml (mean $\pm$ s.e.m., 274 ± 198 ml). The other four produced larger volumes of 1,570–4,470 ml.

Gas analyses showed a significant inverse relationship between the partial pressures of oxygen (p_{O_2}) and carbon dioxide (p_{CO_2}) ($r = -0.837$, $P < 0.001$, $N = 28$, Fig. 2b). p_{O_2} ranged from 0.29–19.9 kPa (1.4–98.2% of the ambient air), whereas the p_{CO_2} varied from 0.36 to 11.8 kPa (11.3–370 times the aerial CO_2). Neither p_{O_2} nor p_{CO_2} was correlated with gas volume. The colour and texture of mud pellets ejected from the burrow by the fish, as well as tracks in the mud around it, indicated the extent of recent



Figure 1 *Periophthalmodon schlosseri* resting in the surface water pool of a burrow. The fish on the left has just emerged from the burrow and is gulping air.

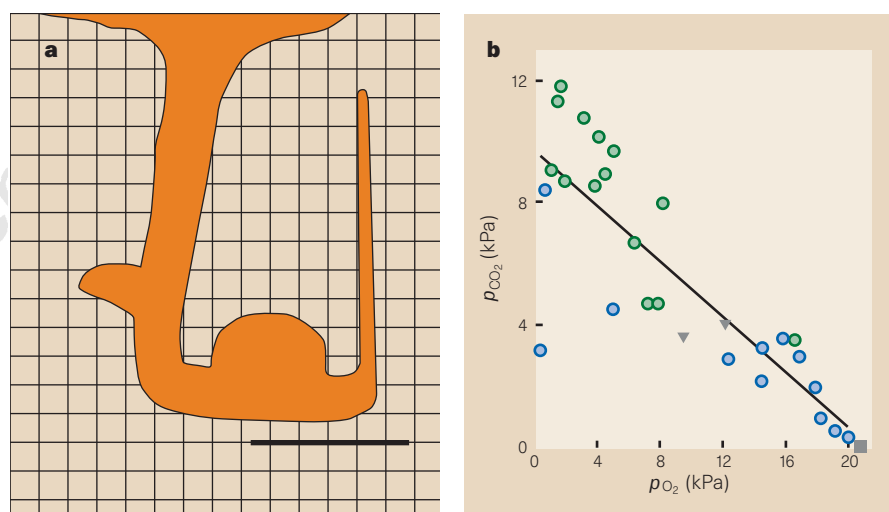


Figure 2 Mudskipper burrow characteristics. **a**, Architecture of the *P. schlosseri* burrow. The drawing is based on a cast of one burrow and direct measurements of another on an estuary of the Burung River (Penang, Malaysia). The *P. schlosseri* burrows are filled with severely hypoxic water. Burrow gas is present in the horizontal tunnel. The shape of a burrow may be highly variable. Scale bar, 30 cm. **b**, Relationship between the p_{O_2} and p_{CO_2} of gas collected from mudflat burrows of *P. schlosseri* in the Burung River estuary (Penang, Malaysia) in August to September 1995 (filled circles), on a tributary of the Selangor River near Kuala Selangor (mainland Malaysia) in November 1996 (open circles), and in the Pinang River estuary (Penang, Malaysia) in December 1996 (filled triangles). Each symbol for burrow gas data represents an average value of duplicate or triplicate determinations of gas samples obtained from one burrow; 28 burrows were studied. Gas was analysed using an O_2 electrode and an infrared CO_2 analyser. Filled square represents air.

activity within a burrow by a fish. It appears that burrows having recent activity contained gases with the higher p_{O_2} and lower p_{CO_2} .

Field observations confirmed that *P. schlosseri* transports air into its burrows. Immediately before entering a burrow, the fish inflated its buccopharyngeal cavity with air. This cavity was deflated when the fish emerged. The fish would immediately take another gulp, and often re-enter the bur-

row. The duration of the burrow sojourns varied; some lasted longer than 30 min. Laboratory studies demonstrated that, irrespective of aquatic oxygen level, *P. schlosseri* regularly inflates its buccopharyngeal cavity with air and, as long as air is accessible, rarely uses aquatic gill ventilation (A. I. and N.M. Aguilar, manuscript in preparation).

Our observations of *P. schlosseri* indicate that the behavioural accumulation of air in the burrows provides a significant oxygen

reservoir for burrow-dwelling fish and for developing embryos. Mudskippers commonly deposit their eggs on the ceiling of their burrows and some species make a specialized spawning chamber for this purpose^{2,4}. Studies in progress show that air is present in the spawning chambers of several other species of mudskipper (*Boleophthalmus dussumieri*, *Oxudercus dentatus*, *Periophthalmus chrysospilos* and *Scartelaos histophorus*; T. Takita, manuscript in preparation).

Atsushi Ishimatsu, Yasuhiro Hishida

Marine Research Institute, Nagasaki University,
Nomozaki, Nagasaki 851-05, Japan

e-mail: a-ishima@net.nagasaki-u.ac.jp

Toru Takita

Faculty of Fisheries, Nagasaki University,
Bunkyo, Nagasaki 852, Japan

Takeshi Kanda

Fisheries Research Laboratory, Miyazaki University,
Nobeoka, Miyazaki 889-05, Japan

Shin Oikawa, Tatsusuke Takeda

Faculty of Agriculture, Kyushu University,
Hakozaki, Fukuoka 812-81, Japan

Khoo Khay Huat

School of Biological Sciences,
Universiti Sains Malaysia,
11800 Penang, Malaysia

- Graham, J. B. *Air-breathing Fishes: Evolution, Diversity and Adaptation* (Academic, San Diego, 1997).
- Clayton, D. A. *Oceanogr. Mar. Biol. Annu. Rev.* **31**, 507–577 (1993).
- Atkinson, R. J. A. & Taylor, A. C. *Symp. Zool. Soc. Lond.* **63**, 133–155 (1991).
- Kobayashi, T., Dotsu, Y. & Takita, T. *Bull. Fac. Fish. Nagasaki Univ.* **32**, 27–39 (1971).
- Clayton, D. A. & Vaughan, T. C. *J. Univ. Kuwait* **15**, 115–138 (1988).
- Murdy, E. O. *Freshwat. Mar. Aquarium*, November, 20–23 (1986).
- Murdy, E. O. *Rec. Austr. Mus.* **11** (suppl.) 1–93 (1989).
- Gordon, M. S., Ng, W. W.-S. & Yip, A. Y.-W. *J. Exp. Biol.* **72**, 57–75 (1978).
- El-Ziady, S., Eissa, S. M. & Al-Naqi, Z. *Bull. Fac. Sci. Cairo Univ.* **48**, 113–133 (1979).

Efficient pollination of alpine plants

Several studies have documented low levels of insect diversity, abundance and activity in alpine ecosystems around the world^{1–7}. It has been hypothesized that these factors may limit pollination of alpine plants^{8,9}. We have measured pollination of alpine and foothill populations of *Campanula rotundifolia* by calculating a corrected visitation rate that incorporated information on pollinator abundance, visitation rates, pollen deposition and duration of stigma receptivity. Although pollinator visitation rates were significantly lower in alpine populations, corrected visitation rates showed that more effective pollination, combined with a longer period of stigma receptivity, compensated for this, resulting in comparable

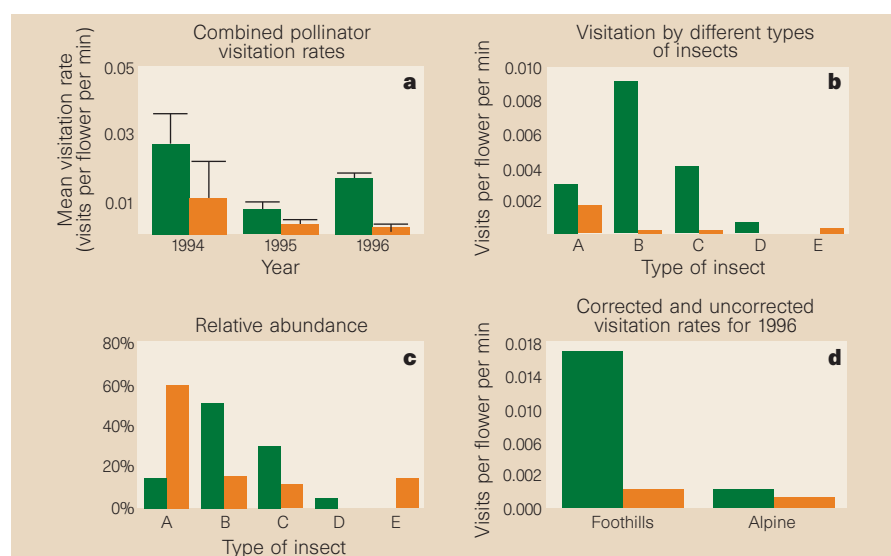


Figure 1 Comparing pollination in alpine and foothill communities. **a**, Mean visitation rates in alpine and foothill populations in 1994, 1995 and 1996. Dark bars represent foothill populations, open bars represent alpine populations. Error bars, 45% confidence intervals. **b**, Mean visitation rates and **c**, relative abundance of the different pollinator categories observed in alpine and foothill populations in 1996. Dark bars, foothill populations; open bars, alpine populations. A, bees; B, medium-sized solitary bees; C, megachilid bees; D, small solitary bees; E, syrphid flies. **d**, Corrected and uncorrected pollinator visitation rates for both alpine and foothill populations in 1996. Dark bars, uncorrected visitation rates; open bars, corrected visitation rates.

levels of pollination in populations from high and low elevations.

We conducted pollination studies from 1994 to 1996 at four low sites (1,800–2,000 m) and five high sites (3,300–3,500 m) in the Rocky Mountains. As in other studies^{1,2,5,6}, we found that low-elevation populations were visited by a more diverse community of pollinators than the high-elevation populations. Pollinator visitation rates were significantly higher in foothill than in alpine populations in all three years of the study ($P < 0.01$ Kruskal–Wallis test) (Fig. 1a). However, visitation rates for specific pollinators differed between elevations. In the foothills, medium-sized solitary bees (*Duforea maura*, *Andrena* sp. and *Colletes* sp.) exhibited the highest visitation rate (0.0109 visits per flower per min), whereas bumblebees (*Bombus* sp.) had the highest visitation rate in the alpine populations (0.0017; $P < 0.05$, Kruskal–Wallis test) (Fig. 1b). Medium-sized solitary bees were also most abundant in the foothills (51%), in contrast to the alpine populations in which bumblebees were the most common visitor (60%) (Fig. 1c). As in other studies^{7,10}, we also found that bumblebees deposited significantly more pollen per visit than any other type of pollinator (ANOVA, $P < 0.001$).

The duration of stigma receptivity also differed significantly between alpine and foothill populations. In 1993, alpine flowers were receptive for 2.31 days, whereas foothill flowers were receptive for 1.46 days ($t = -3.14$, $P < 0.001$). We obtained similar results in 1995 (3.00 days as opposed to 2.21 days, $t = 4.07$, $P < 0.0005$). We used these

data to calculate a corrected visitation rate for alpine and foothill populations (Fig. 1d). All foothill visitation rates were multiplied by $2/3$ to correct for the shorter duration of stigmatic receptivity at low elevations. That quantity was then multiplied by a measure of relative pollen deposition for each pollinator category and its relative abundance at each elevation. In the foothills, the uncorrected visitation rate for all insect types combined was 0.0168 in contrast to a corrected rate of 0.0019. The uncorrected visitation rate in the alpine populations was 0.0023, and the corrected rate was 0.0011.

Although pollinator diversity and activity in populations of *C. rotundifolia* decrease with increasing elevation, our results show that pollination in alpine and foothill populations is more comparable than would be expected from visitation rates and insect diversity alone. This is in contrast to the widespread assumption that restricted insect activity at high elevations limits pollination of alpine plants relative to that at lower elevations. The most important factors contributing to these results are the dominant role of bumblebees as pollinators in alpine populations and the longer duration of stigmatic receptivity in high-elevation plants. The latter point is particularly relevant to recent research that has investigated increased floral longevity as an adaptation to low pollinator visitation rates¹¹.

Robin A. Bingham

Department of Environmental, Population,
and Organismic Biology, University of Colorado,
Boulder, Colorado 80309, USA
e-mail: rbingham@western.edu

Andrea R. Orthner

Biology Department, Oberlin College,
Oberlin, Ohio 44074, USA

1. Inoué, D. W. & Pyke, G. H. *Aust. J. Ecol.* **13**, 191–210 (1988).
2. Arroyo, M. T. K., Primack, R. & Armesto, J. *Am. J. Bot.* **69**, 82–97 (1982).
3. Arroyo, M. T. K., Armesto, J. J. & Primack, R. B. *Plant Syst. Evol.* **149**, 187–203 (1985).
4. Müller, H. *Nature* **21**, 275 (1880).
5. Moldenke, A. R. & Lincoln, P. G. *Phytologia* **42**, 349–379 (1979).
6. Kearns, C. A. *Am. Midland Nat.* **127**, 172–182 (1992).
7. Kearns, C. A. & Inoué, D. W. *Am. J. Bot.* **81**, 1091–1095 (1994).
8. Billings, W. D. in *Arctic and Alpine Environments* (eds Ives, J. D. & Barry, R. G.) 403–438 (Methuen, London, 1974).
9. Löve, A. *Publ. McGill Univ. Mus.* **1**, 82–95 (1959).
10. Fishbein, M. & Venable, D. L. *Ecology* **77**, 1061–1073 (1996).
11. Ashman, T.-L. & Schoen, D. J. *Nature* **371**, 788–790 (1994).

How molecular motors work in muscle

Much of Howard's Review Article¹ concerns the results of experiments with single myosin molecules and actin filaments. It was a huge surprise when reports of such experiments first appeared (see, for example, ref. 2), and much is being learnt from them that cannot be deduced from experiments on whole muscle fibres, whether intact or after removal of the membrane. But single-molecule experiments do not yet approach the time resolution or the freedom from brownian noise that are easily attainable on larger assemblies of myosin and actin filaments, and their interpretation is subject to many uncertainties — due, for instance, to compliance in the actin filaments and in their attachments to beads or other force-measuring components, and the attachment of myosin molecules or fragments to the base. No doubt the time course of the working stroke of a single myosin head will one day be recorded, but until that is achieved the results of experiments on whole fibres and myofibrils deserve more careful attention than has been given to them by Howard.

Many of Howard's statements can be challenged. Here I will mention only three, of particular relevance to my own work. First, Howard uses a value of 4 nm for the "working distance" of a myosin molecule, defined as the distance over which a myosin head remains attached during a single interaction with an actin filament. But he uses the same value of 4 nm for the "working stroke", the distance over which an attached crossbridge exerts positive force. In rapid shortening, the former is presumably greater than the latter as crossbridges remain attached for a short time after the force they exert has fallen to zero. Howard justifies the use of this value on the grounds that it is the amount of sudden shortening per half-sarcomere needed to bring tension

to zero from its value in an isometric tetanus of an isolated frog muscle fibre (Fig. 26 of ref. 3).

However, this quantity is usually interpreted as the average amount by which compliant structures in the half-sarcomere had been strained by the active tension before application of the shortening step. The working stroke is represented in such experiments by the rapid recovery of much of the tension that was present before the shortening step. The slack in a fibre caused by sudden shortening of 13 nm per half-sarcomere is the maximum taken up by this quick recovery process (Fig. 13 of ref. 3). In simulations of the response to a shortening step, this maximum is about equal to the assumed working stroke plus the contribution from linear compliance, whether it is assumed that the working stroke is performed by crossbridges that were already attached before the step (as in my own⁴ and many other simulations) or by fresh crossbridges that attach very rapidly after the step. In the latter case, if a stroke of 4 nm is assumed the model cannot account for the 13 nm of shortening observed (Fig. 7 of ref. 5).

Second, Howard points out correctly that "the discovery that the actin filaments contribute about half the compliance of muscle [during isometric contraction] means that the stiffness is not proportional to the number of attached heads" and concludes that "a low duty ratio is therefore not inconsistent with the stiffness measurements" (p. 564 of ref. 1).

The stiffness of an isolated frog fibre contracting under zero load is about one-third of its stiffness during isometric contraction⁶, implying 20% of heads attached if it is assumed that all heads are attached in rigor. This is very different from the 1% claimed by Howard.

Last, Howard writes of "a paradox" — that the amount of filament sliding that occurs in the time required by each myosin molecule to hydrolyse one ATP is much larger than the "working distance", referring to two papers (refs 7 and 8 here). He claims to resolve this "paradox" by supposing that much of the sliding takes place while the myosin is detached from actin.

But for many years (for example, ref. 9) it has been supposed that each myosin head acts intermittently and that continuous sliding is brought about by asynchronous action of many myosin molecules, although estimates of the sliding distance per ATP hydrolysed have varied. The controversy raised by refs 7 and 8 was different: those papers report that when a myosin head interacts with an actin filament during rapid shortening, it *remains attached* for a distance of 60 nm (ref. 7) or 40 nm (ref. 8). Staying attached for such large distances remains difficult to explain, and Howard's

"resolution" of the paradox is irrelevant because his central postulate is that a myosin head remains attached for only 4 nm, much smaller than 40 or 60 nm.

Andrew Huxley

Trinity College,
Cambridge CB2 1TQ, UK

Howard replies — Huxley points out that there is a discrepancy between recent mechanical recordings from single myosin molecules *in vitro* and earlier mechanical measurements from muscle fibres *in vivo*: the single-molecule recordings indicate a working distance and maximum duty ratio of 4–6 nm and 10–20%, respectively, only about half the values deduced from the fibre studies. Although it is possible that technical limitations, for example misorientated heads, make the single-molecule recordings suspect, I argued in my Review Article¹ that the discrepancy might be only an apparent one. In particular, the smaller working distance does conform to many of the results from muscle, and the lower duty ratio can even provide a more satisfactory explanation for some of the experiments with muscle fibres. Huxley challenges this view by discussing three additional observations from fibres. But I believe that these too are mainly consistent with the smaller working stroke and duty ratio.

First, Huxley argues that the working distance corresponds to the approximately 13-nm range of displacements over which muscle tension quickly recovers after a sudden shortening³. This is true if most of the myosin heads (crossbridges) are bound to the actin filament — that is, if the duty ratio is large.

However, one of the features of the low-duty-ratio model is that the relative movement of the actin and myosin filaments will bring unattached myosin heads into striking distance of new actin-binding sites. In this way the large range of displacements can be explained by the rapid attachment of fresh crossbridges (and detachment of the initially attached ones)⁵. In this view, the quick-recovery distance is not necessarily limited by the working stroke, and in principle the force might even recover over a substantial fraction of the periodicity of the helical actin repeat (see discussion of Fig. 7 in ref. 5).

Huxley's second two points concern myosin's duty ratio. In the earlier work on muscle fibres, the duty ratio was calculated as the ratio of the stiffness of active muscle to that of rigor muscle (in the latter case it is known that all the myosin heads are attached to the actin filaments¹⁰). This calculation yields a duty ratio significantly higher than that estimated from the more recent single-molecule, biochemical and other structural data that I summarized¹.

However, a key untested assumption is

that all the crossbridges in a rigor muscle contribute to the stiffness. If this is not so — for example if many are bound in 'slack' states — the calculated duty ratio would be lower and the discrepancy might vanish. If the duty ratio is indeed small, the experiments of Higuchi and Goldman⁸ are consistent with a small working distance. The experiment of Yanagida *et al.*¹¹ remains difficult to explain.

Joe Howard

Department of Physiology and Biophysics,
University of Washington, Seattle,
Washington 98195, USA
e-mail: johoward@uwashington.edu

- Howard, J. *Nature* **389**, 561–567 (1997).
- Finer, J. T., Simmons, R. M. & Spudich, J. A. *Nature* **368**, 113–119 (1994).
- Ford, L. E., Huxley, A. F. & Simmons, R. M. *J. Physiol.* **269**, 441–515 (1977).
- Huxley, A. F. & Tideswell, S. J. *Musc. Res. Cell Motil.* **17**, 507–511 (1996).
- Eisenberg, E., Hill, T. L. & Chen, Y. *Biophys. J.* **29**, 195–227 (1980).
- Ford, L. E., Huxley, A. F. & Simmons, R. M. *J. Physiol.* **361**, 131–150 (1985).
- Yanagida, T., Arata, T. & Oosawa, F. *Nature* **316**, 366–369 (1985).
- Higuchi, H. & Goldman, Y. E. *Nature* **352**, 352–354 (1991).
- Huxley, A. F. *Prog. Biophys. Biophys. Chem.* **7**, 255–318 (1957).
- Lovell, S. J., Knight, P. J. & Harrington, W. F. *Nature* **293**, 664–666 (1981).
- Yanagida, T., Arata, T. & Oosawa, F. *Nature* **352**, 352–354 (1985).

A new classification for HIV-1

The phenotype of HIV-1 isolates is defined by the cells in which they replicate *in vitro*, but these phenotypes can change *in vivo* with profound implications for viral transmission, pathogenesis and disease progression. Here we propose a new classification system based on co-receptor use, providing a more accurate description of viral phenotype than the present imprecise and often misleading classification schemes.

At present, three classification systems are in use. The first defines primary isolates as macrophage (M)-tropic or T-cell-line (T)-tropic. However, this system disguises the fact that all primary isolates replicate in activated, primary CD4⁺ T-lymphocytes. The second system categorizes isolates as being either syncytium-inducing or non-syncytium-inducing (NSI) on the basis of whether they form syncytia in MT-2 cells, which express CXCR4 but not CCR5 (ref. 1). However, NSI viruses can readily form syncytia with CCR5-positive cells. The third system defines viruses as either slow/low (SL) or rapid/high (RH) depending on their growth kinetics in culture². These classifications are often used interchangeably, but they are not synonymous.

The identification of some chemokine

receptors as having critical roles in the cellular entry of HIV-1 allows us to develop a more precise system for identifying the phenotypic properties of virus strains. A major determinant of HIV-1 tropism (phenotype) lies at the level of virus entry into target cells, which in turn is governed by the expression of co-receptors in conjunction with CD4 (refs 3–5): either CCR5 or CXCR4, or both. CXCR4 use is a defining feature of viruses that form syncytia in T-cell lines; use of CCR5 is a property of NSI, M-tropic viruses; and many T-tropic primary isolates can use both co-receptors^{3–5}. A nomenclature based on the co-receptor used would thus provide a precise molecular designation of a given isolate that largely explains its phenotype.

We propose that isolates that use CCR5 but not CXCR4 be termed R5 viruses, that isolates using CXCR4 but not CCR5 be designated X4 viruses, and isolates able to use both co-receptors with comparable efficiency be called R5X4. Whether an X4 or R5X4 virus is a cell-line-adapted isolate should also be specified.

Under this system, R5 viruses are the strains most commonly transmitted sexually, consistent with the high resistance of individuals lacking CCR5 to infection^{6,7}. After about five years, viruses evolve in about 50% of patients that are able to use CXCR4, with or without concurrent use of CCR5 (refs 8,9). These viruses would now be called R5X4 and X4 viruses, respectively. Isolates passaged through a permanent T-cell line should be called T-cell line-adapted (TCLA) X4 or R5X4 viruses, and a similar qualifier can be used for viruses adapted to growth on other cells.

This nomenclature takes note of the ability of an isolate to use the major co-receptors, but does not specify whether the isolate can replicate in a particular target cell. The nuances of co-receptor usage in specific contexts are beyond a simple classification system and should be specified by authors if there are perceived ambiguities. That a virus can use a particular co-receptor in transfected cells does not mean that this virus uses the same co-receptor in a more physiological context. The efficiency with which different co-receptors are used by some strains is likely to vary between assay systems; again, authors should clearly explain the limitations of their results, and the significance of extremely low efficiency co-receptor usage should not be over-interpreted. This classification system can also be expanded to take note of other co-receptors if their use by an isolate proves to be a major determinant of tropism; for example, whether isolates use CCR3 and/or CCR5 to enter microglia could define them as R3, R5 or R3-R5 isolates¹⁰.

Our classification system is uncomplicated, is intuitive to those familiar with the

usage of the CCR5 and CXCR4 co-receptors by HIV-1, and removes the inaccuracies and confusion associated with the present systems. It is flexible and open to expansion to accommodate emerging knowledge of co-receptor usage, and can encompass HIV-2 and SIV strains, which also use CCR5 and other co-receptors for entry^{3–5}.

A record of co-receptor use by particular HIV-1, HIV-2 and SIV strains will be maintained by the Los Alamos National Laboratory Sequence Database, together with an expanded version of this article.

E. A. Berger

Laboratory of Viral Diseases, NIAID,
National Institutes of Health, Building 4,
Room 236, Bethesda, Maryland 20892, USA

R. W. Doms

Department of Pathology and Laboratory Medicine,
University of Pennsylvania, Philadelphia,
Pennsylvania 19104, USA

E.-M. Fenyo

EC Concerted Action on HIV Variation,
Microbiology and Tumorbiology Center,
Karolinska Institute, 171 77, Stockholm, Sweden

B. T. M. Korber

Theoretical Biology and Biophysics, Los Alamos
National Laboratory, Los Alamos,
New Mexico 87545, USA

D. R. Littman

Howard Hughes Medical Institute,
New York University Medical Center, New York,
New York 10016, USA

J. P. Moore

The Aaron Diamond AIDS Research Center,
The Rockefeller University, 455 First Avenue,
New York, New York 10016, USA

Q. J. Sattentau

Centre d'Immunologie de Marseille-Luminy,
Case 906, 13288 Marseille, France

H. Schuitemaker

Department of Clinical Viro Immunology,
Central Laboratory of the Netherlands Red Cross
Blood Transfusion Service and Laboratory for
Experimental and Clinical Immunology,
University of Amsterdam, Amsterdam,
The Netherlands

J. Sodroski

Department of Cancer Immunology/AIDS,
Dana-Farber Cancer Institute, 44 Binney Street,
Boston, Massachusetts 02115, USA

R. A. Weiss

Chester Beatty Laboratories,
The Institute of Cancer Research, 237 Fulham
Road,
London SW3 6JB, UK

- Schuitemaker, H. *et al. J. Virol.* **66**, 1354–1360 (1992).
- Fenyo, E. *et al. J. Virol.* **62**, 4414–4419 (1988).
- Berger, E. A. *AIDS* **11** (suppl. A), S3–S16 (1997).
- Moore, J. P., Trkola, A. & Dragic, T. *Curr. Opin. Immunol.* **9**, 551–562 (1997).
- Doms, R. W. & Peiper, S. C. (1997) *Virology* **235**, 179–190.
- Liu, R. *et al. Cell* **86**, 367–377 (1996).
- Samson, M. *et al. Nature* **382**, 722–725 (1996).
- Richman, D. D. & Bozzette, S. A. *J. Infect. Dis.* **169**, 968–974 (1994).
- Connor, R. I. & Ho, D. D. *J. Virol.* **68**, 4400–4408 (1994).
- He, J. *et al. Nature* **385**, 645–649 (1997).

The history of healing laid bare

The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present

by Roy Porter

HarperCollins: 1997. Pp. 831. £24.99.

To be published in the United States by Norton in April at \$35

Alfred W. Crosby

I intend to keep this book within reach for years to come. It matches the older histories of medicine in quality, and supersedes them in its coverage of many areas, especially the current situation with AIDS and rocketing medical costs. And its bibliography is truly useful, being selective rather than pseudo-comprehensive and stupefying.

This is a history specifically of medicine, not of myriad related matters such as medical beliefs, health, quackery and demography, although these are granted space whenever they are so important that it would be absurd to exclude them, for example the population crashes triggered by the European invasions of America and Africa. Many pages are devoted to institutional history, such as the growth of hospitals, medical schools and research institutions, to the professionalization of healing, to medical societies and systems of certification, and to the intervention of the state in medical matters.

Such tomes as this should be factually flawless but never are. Sorry, Professor Porter, 30 were not executed for witchcraft at Salem, Massachusetts, though 20 were. And 40,000 did not die of measles in Hawaii in 1848, but more like 10,000. Other critics with different funds of specialist knowledge will probably find other errors to carp about. Writing a history book without mistakes is like living a life without sin: worth the effort, but impossible.

A book of this scope should be judged not on picky detail but on how well it achieves what the author intended. The subtitle of this one is *A Medical History of Humanity* (I'll get to the main title in a bit). Of all humanity? Nonsense. As it is, the book stretches itself as thin as a good history of Western medicine will go without tearing, with token chapters on Indian and Chinese medicine. It is an example of what multiculturalists call 'the West and the rest' school of historiography.

We might be able to pay proper respect to non-Western medical traditions and, on the same page, grant that antibiotics are better for plague than moxibustion if we give up our habit of dividing the history of medicine geographically and divide it chronologically. From the last Ice Age to about 200 or so years ago, nothing much happened in the way of curing infections and dealing with rebellious appendixes. Then something did happen, the 'great acceleration'. It happened demographi-



A Barber Surgeon Tending a Peasant's Foot by Isaac Koedijck, 1650, from the cover of Porter's book.

cally, industrially, technologically, in science in general, and in medicine too. It started in the West, but spread with unprecedented swiftness. One of the top bacteriologists in the world at the end of the nineteenth century was Shibasaburo Kitasato, and at the end of the twentieth century the Swedes are giving Nobel prizes for physiology or medicine to people who really don't look like Swedes at all.

I recommend that we should clump medical history before the eighteenth century into one category with two divisions: folk medicine and rational medicine. The first produced much of value, such as smallpox variolation, but could not develop any further than it did because it was a matter of magic or pure and stifling empiricism.

Rational medicine (diagnosis and treatment on the basis of careful observation and logical analysis) did not cure many, but neither did it kill many, and it created a philosophical base for scientific medicine. Some of its practitioners were Greek, some Ayurvedic, some Chinese, some Arabic. They spoke of humours, of health being a matter of balances, of matching the cycles of the body with those of the cosmos; they all strike me as sounding much more like each other than like Western-trained physicians since the therapeutically nihilist, autopsy-centred Paris tradition of the early 1800s.

Arranging medical history chronologically saves us from ranking Hippocrates as somehow better than the best of today's Asian physicians, and leaves us with the real mys-

tery: why did modern medicine burst onto the scene when and where it did?

Porter pays homage to that explosion by devoting the entire middle of his book to Western medicine in the nineteenth century. This is textbook history at its best and worst. All the names you remember (or should remember) are there: Liebig, Bernard, Lister, Pasteur, Koch, Virchow, Semmelweis, Nightingale and on and on, each with date, place of birth and achievements. Every sentence can be justified, but after 50 or so pages it feels like trying to understand New York city by leafing through its telephone directory.

The chapters on the specifics of research science, public health, bacteriology, psychiatry, immunology and so on in the nineteenth and twentieth centuries make up the bulk of the book, and for many students will be the most valuable. They would have made better reading if Porter had cranked himself up out of the ocean of details to a higher level of abstraction, but that would make a different kind of book from what he intended.

In the final chapters Porter battles successfully against the avalanche of details about the course of medicine in our century, its soaring triumphs, mountainous costs, and decline in public esteem. These are the best chapters because he matches strong opinions to his mastery of the facts. His final sentences sum up medicine's history and its current dilemma. For millennia it attempted little, and failed at that. Today it has accomplished what would have been called near miracles as

recently as 1900. But the public has expectations beyond even this. His book ends with the thought that "medicine will have to redefine its limits even as it extends its capacities".

The book's main title is *The Greatest Benefit to Mankind*. There are solid and obvious arguments to support that claim for medicine. But the past is full of bitter ironies, and so may be the future. If the antibiotic revolution produces more antibiotic-resistant pathogens than it counters — or if medicine continues to drive death rates down and live birth rates up, swamping the planet with people — will our descendants still think of it as humanity's greatest benefit? □

Alfred W. Crosby is at the Renvall Institute, University of Helsinki, PO Box 59, 00014 Finland.

Industrial revolutions

The One Best Way: Frederick Winslow Taylor and the Enigma of Efficiency

by Robert Kanigel

Viking/Little, Brown: 1997. Pp. 675. \$34.95, £20

Endless Novelty: Specialty Production and American Industrialization

by Philip Scranton

Princeton University Press: 1997. Pp. 415. \$39.50, £29.95

David E. Nye

How should we tell the story of the second industrial revolution, which began in the United States in the late nineteenth century and by the 1920s resulted in massive productivity increases, high wages and the growth of consumer society? Philip Scranton and Robert Kanigel offer quite different answers to this question.

Kanigel tells the more familiar story in *The One Best Way*, a biography of Frederick Winslow Taylor, a founder of scientific management. Taylor believed that industrial work could be reduced to scientific principles. There was "one best way" to do any job, and once this was discovered by an expert, stopwatch in hand, workers were to conform to the system. Taylor would teach the most efficient movements and invent the ideal tools for even the simplest task. Workers who cooperated were more productive, and, in return, managers were expected to pay them more.

Educated in European schools and at the Phillips Exeter Academy in New Hampshire before he chose a machine shop over Harvard, Taylor was a persuasive speaker and writer. After 1908 he lectured annually at Harvard Business School, and his paper trail is more enticing than that of the managers and workers who opposed him. In lively prose sprinkled with anecdotes, Kanigel describes Taylor's privileged boyhood in Philadelphia, his choice of practical work over gentility, and



Boston lawyer Louis Brandeis was the instrument of Taylor's rise to fame. In a case in which Brandeis opposed some US railroad companies' application for an increase in freight tariffs, he championed the system of 'scientific management', lecturing the companies on how they could save millions of dollars.

his rise from the shop-floor to champion rationality and efficiency. His principles, Kanigel argues, reshaped contemporary life in a thousand subtle ways. "Taylor's impact was like that of Darwin, Marx, and Freud", for each brought an analytical "cast of mind to an unruly, seemingly intractable problem".

Kanigel is better on Taylor's ancestry, education and career than on his place in industrial history. His portrayal of Taylor is more sympathetic than (and three times as long as) Daniel Nelson's 1980 biography, and is perhaps overly detailed in recounting Taylor's private life. Kanigel's Taylor is not so much arrogant as self-assured, owing to a good family background. His eye problems and headaches, which prompted him to abandon formal schooling, are not psychosomatic but caused by poorly understood astigmatism. Frequent conflicts with both workers and managers are caused less by insensitivity than by a theatrical personality. Overall, he appears less an eccentric than a bull-headed individualist.

Taylor's ideas had enormous resonance beyond the factory in home economics, education, religion and popular culture. His *Princi-*

ples of Scientific Management sold widely in French, German and Russian translations, and was praised by Lenin. Taylor was lionized by Louis Brandeis and later by many who were prominent in the New Deal. Yet labour leaders loathed his ideas, which were subjected to some rather hostile congressional hearings in 1912.

Kanigel argues that Taylor's system largely anticipated the assembly line. This is nonsense. Taylor made shovelling more efficient; Ford used electrified machines to eliminate shovelling. Taylor retained piece-work, a concept that makes no sense on the assembly line, where all share the same pace. Taylor maximized efficiency in existing production; Ford transformed the means of production. Taylor saved time; Ford sped up time.

In 1914, after Taylor imposed his system on the Packard Motor Company, its 4,525 workers produced 2,984 cars a year, less than one per man. This was still artisanal production of a luxury vehicle. Ten years earlier, well before the assembly line, Ford was already producing 12 cars for every worker. By 1914, 13,000 Ford employees, using conveyor belts, electric cranes and 15,000 specialized machine tools, manufactured 260,000 cars, or

20 per worker. Ford out-performed Taylor by more than 2,000 per cent.

In contrast to either of these stories of mass production, Scranton's *Endless Novelty* develops a more nuanced and ultimately more convincing account of the second industrial revolution. Although his argument is more complex than Kanigel's, nonspecialists will find Scranton's case studies accessible.

Rather than focusing on large corporations such as Standard Oil, Ford and DuPont as the essence of advanced capitalism, Scranton examines smaller companies with customized and small batch production. Far from being a backwater destined to be converted to 'Taylorism' or assembly lines, such companies were as important as mass production to an advanced economy. Companies fabricating elevators, turbines, switchboards and precision instruments did not mimic the mass producers. They tended to remain family companies or closely held corporations. They did not de-skill labour or subject it to routines; rather, they prized skilled workmen who could respond flexibly to variations in demand.

Such workers fashioned the endless novelty that was the hallmark of the consumer society, and such companies were innovative and profitable. They had little use for standardization but explored various productive systems. A system "could increase reliability, reduce errors, speed batch throughput... but if it went so far as to freeze flexibility and innovation, it would become a burden and a constraint". In the marketplace of systems, Taylor's prescriptions were useful for the repetitive production of identical things, but not for the batch production of carpets, furniture, jewellery, cutlery, hats and ready-to-wear clothing. Then, as now, middle-class consumers wanted endless novelties, and large profits accrued to the companies with skilled workers and flexible modes of production that could supply them.

Scranton argues from case studies of a dozen representative industries between the American Civil War and the late 1920s. The six producer- and six consumer-oriented companies he examines in detail include manufacturers of jewellery, furniture, carpets, textiles, clothing, steel, steam engines, saws and locomotives.

Scranton's great strength, making summary difficult, is that he does not lump these companies into a single category but rather points out the subtle differences inherent in their businesses. He also locates them in relation to familiar, large corporations such as General Electric, which by 1925 included both the production of specialized goods (such as turbines and switchboards) and the mass production of appliances. Scranton also builds geographical diversity into his argument, taking the reader on tours of Philadelphia (locomotives, textiles, carpets, ships and tools), New York (printing and clothing), Provi-

dence (steam engines, jewellery and sewing machines), Chicago (Pullman railroad cars, metal-working and furniture), Cincinnati (machine tools), and other industrial cities.

These case studies represent sizeable sectors of the economy. In 1890, foundries and machine shops had 248,000 workers, printing and publishing 165,000. The production of specialized goods grew just as fast as mass production, each of which in 1909 represented a third of national output. Of the two, the production of specialized goods added more value and employed more workers. By 1923, value added by such companies had tripled, and they still accounted for a third of industrial employment.

These companies matched the explosive growth of mass production, but by different means. Instead of the "visible hand" of visionary managers guided by scientific experts, at the centre stand skilled workers, technologically adept owners, and competing systematizers, of whom Taylor was but one. Instead of "one best way", there was flexibility. Instead of standardization of production at capital-intensive plants, there was diverse production at labour-intensive, mid-sized factories which proved technologically innovative.

In this compact and meticulously researched study, Scranton shows that scientific management could not have been applied to more than a third of the economy, and that mass production alone cannot explain the second industrial revolution. □

David E. Nye is at the Center for American Studies, Odense University, Campusvej 55, Odense 5230 M, Denmark.

At the sharp end

Projectile Technology

edited by Heidi Knecht
Plenum: 1997. Pp. 408. \$59.50, £43.27

John A. J. Gowlett

A chimpanzee can lob a brick quite effectively. A shared ancestry of long, strong arms, mobile shoulder joints and deft hands suggests that hominids too might have a long history of throwing things.

The oldest known wooden spear dates back some 400,000 years, and somewhere along the line people started to incorporate stone tips into what became composite weapons. Recent times have provided a worldwide plethora of projectiles, including spears and spear-throwers, bows, arrows and darts. In between these outline points there is a whole world for archaeology to explore.

Although an evolutionary perspective is inviting, the place to start is the present, because there is a vast amount of evidence, and a great range of weaponry, varied for reasons that are not always obvious. It seems appropriate that an American archaeologist should set out to bring together recent

research across this field, because carefully shaped projectile points, striking to the eye, have been an important part of tool kits across the Americas for at least the past 12,000 years.

Heidi Knecht introduces a set of papers written by 19 contributors, including herself, all concentrating on projectile technology, and they pepper their target. For anybody who is intrigued by elementary — if not actually simple — technologies, much of this reading is a delight, even if technical detail occasionally takes over.

Authors such as John Shea who write about the earlier Palaeolithic manage to provide a broad evolutionary perspective. But modern humans have made most of the projectiles ever used, through at least the past 40,000 years, and it is fair that the book should be mainly about their products. Some of these are complicated and seem unlikely, such as the iron-tipped arrows of the Agta people in the Philippines which can look like the skeleton of a fish.

Compromise is the reality for people hunting animals, as their decisions involve how to approach prey as well as what technology to use. It is as easy to miss with a spear as with a bow. As emphasized by Robert Hitchcock and Peter Bleed, "...hunting proceeds through a number of steps, each of which carries the possibility of failure". Most shots are delivered over a maximum of 25 metres, often less than half that. 'Optimality' involves choices about 'durability', 'maintainability' and 'robusticity', as the discussions show.

To understand what is going on there is a great need for some actualistic reference: experiment on the one hand, and ethnoarchaeological observation on the other. What were the relative advantages of spear-throwers and bows for Upper Palaeolithic peoples? How many uses could a Mesolithic microlith have? It is encouraging to see the variables whittled down, and real progress made in the interpretations. There is not quite consensus, but spears make a far greater impact than arrows, and poison helps the latter do their damage.

A strength of the book is its wide range, ensured by having contributors from several traditions including American, French and British. The section on experimental work is



substantial, extending to archaeological comparisons. The ethno-archaeological chapters are among the most readable, dealing with the well-known hunters of Africa and their present problems as well as their technology. Hunting practices of Venezuela and the Philippines also feature. World coverage cannot be comprehensive, but fortunately some authors give broad surveys. Pierre Cattelain, for example, discusses Australian technologies in his study of the Upper Palaeolithic, and provides numerous references.

Both editor and contributors have done well, despite a slightly wooden format, now too common, and probably imposed by publishers or a growing convention. Editors should have their say — but just once. When they write introductions to each and every section, and even to the concluding retrospective chapter, it implies that formula is more important than effective use of space. A concluding chapter is useful, however, and Margaret Nelson makes many pertinent comments, dealing *en passant* with sexism: if studies of projectiles and hunting tended to be by men about men, excellent studies here and elsewhere show that there is no good reason for that.

One has to be enthusiastic about the scholarly value of a book so packed with information and case studies. Spears and arrowheads, exotic and mundane, are all here, together with many analyses of hunting trips and decision paths, which are invaluable in showing how hunter-gatherers spend their time, and how they get their returns. □

John A. J. Gowlett is in the Department of Archaeology, University of Liverpool, Liverpool L69 3BX, UK.

Technological heaven

The Religion of Technology: The Divinity of Man and the Spirit of Invention

by David F. Noble

Knopf: 1997. Pp. 273. \$26

Howard P. Segal

The centuries-old belief in unceasing technological progress leading to Utopia has often been equated with the religious belief in salvation or perfection through divine intervention. Similarly, engineers and inventors have repeatedly been characterized as secular deities in their reputed ability to transform the world.

Had David F. Noble contented himself with merely updating this ideology as it applies to endeavours since the Second World War in nuclear weapons, space flight, artificial intelligence and genetic engineering, he would have had an interesting but hardly original book. Instead, he has also provided a ground-

breaking account of the religious, especially Christian, roots of Western technology from medieval times to the present day.

Rejecting the still common assumption that religion and technology, like religion and science, have always been antagonistic, Noble shows how major Western technological advances, like major Western scientific discoveries, have routinely been invested with religious significance. Many Western technological pioneers have linked their achievements with the recovery of human divinity, lost after Adam's fall.

For Noble, the 'religion of technology' is not a mere metaphor but a fact of life. What, for obscure reasons, began in the early Middle Ages as the linkage by Benedictine monks of redemption with the 'useful arts' — windmills, watermills, clocks, metal forges and ploughs, for example — steadily extended to most successive Western technologies. Given this revisionist ideological framework, it is hardly surprising to see twentieth-century technological élites increasingly eager to transcend earthly boundaries through space flight and to exercise God-like powers of genetic engineering and artificial intelligence.

These technological élites have been and remain overwhelmingly male, aggressively masculine in their culture and values. As important to Noble's revisionist argument as the bonding of Christian theology with technological advance is the exclusion of women from these technological crusades. Indeed, the ultimate dream of many of the visionaries he discusses is the creation through genetic engineering of a womanless world, an "Eden before Eve", as Noble put it in his 1992 book *A World Without Women: The Christian Clerical Culture of Western Science*.

The Religion of Technology is a sequel to *A World Without Women*, which traced the continuing inferior status of women in most scientific and technical fields back to the medieval Latin Church. After centuries of considerable gender equality, the Church began excluding women from the monasteries, scientific societies and early universities from which modern Western science sprang. Noble places excessive blame on Christianity, which was hardly unique among world religions and cultures in relegating women to second-class status, at best the curious objects of male 'scientific' attention.

But Noble's illumination of relative gender equality within pre-medieval Christianity reinforces the fundamental theme of all his writings: science and technology are always the products of human endeavours and decisions, not omnipotent forces. As Noble has repeatedly reminded us, those who invoke the omnipotence of science and technology invariably have vested interests in actually controlling scientific and technological applications. Consequently, notwithstanding the theological underpinnings detailed in both of Noble's works, nothing was inevitable — or

'natural' — about either the separate spheres created for men and women or the millenarian thrust of Western science and technology.

Like *A World Without Women*, *The Religion of Technology* relies heavily on secondary works, as Noble, a distinguished scholar of American technology, readily admits in the former volume but not in the latter. Perhaps to reassure readers that his facts (if not necessarily his controversial interpretations) are firmly grounded in research, Noble in both books quotes excessively from scholars of fields in which he is not expert. Yet his powerful overarching arguments and his own clear, compelling prose overcome these limitations.

Nevertheless, one wonders about the pervasiveness of the *Religion of Technology* in the light of the growing scepticism in recent decades about technological progress in general. This scepticism, which reflects inevitable disillusionment when technological utopias do not come about, is partly in reaction to the very technological advances Noble discusses. As Noble recognizes, the technological élites he condemns for their arrogance care little about ordinary people (males and females alike) and would gladly leave them behind on Earth as they pursue glorious space flights and space colonies. They would reproduce only geniuses like themselves through genetic engineering or create computerized virtual minds lasting indefinitely. In fact, Noble connects their indifference to the repeated failure of modern technology to meet basic human needs for all as opposed to fulfilling the self-centred dreams of a few.

Still, the mundane daily concerns of millions everywhere about technology's profoundly mixed blessings — as detailed in Edward Tenner's *Why Things Bite Back: New Technology and the Revenge Effect* (Knopf/Fourth Estate, 1996; for a review see *Nature* **382**, 504; 1996) — make the dreams of these engineers and inventors more *passé* than avant-garde, more pathetically naive than enviably sophisticated. The transformations those technological élites deem almost inevitable are surely far from certain.

In addition, and partly because of the commonplace frustrations and disappointments illuminated by Tenner, there no longer exists in the countries in which these visionaries reside the grass-roots enthusiasm for space programmes, genetic engineering, and other millenarian enterprises that governments and corporations could once take for granted in allocating funds. Paradoxically, then, widespread technological scepticism may offer the world more hope for genuine technological — and social — progress than Noble allows.

Nevertheless, *The Religion of Technology* is a most significant work that deserves a wide readership. It is a worthy successor to Lewis Mumford's critical histories of technology. □ Howard P. Segal is in the Department of History, University of Maine, 5774 Stevens Hall, Orono, Maine 04469-5774, USA.

Constraints on cortical and thalamic projections: the no-strong-loops hypothesis

Francis Crick* & Christof Koch†

* The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

† Computation and Neural Systems Program, 139-74, California Institute of Technology, Pasadena, California 91125, USA

The many distinct cortical areas of the macaque monkey visual system can be arranged hierarchically, but not in a unique way. We suggest that the connections between these cortical areas never form strong, directed loops. For connections between the visual cortex and particular thalamic nuclei, we predict that certain types of connections will not be found. If strong, directed loops were to exist, we suggest that the cortex would go into uncontrolled oscillations.

The macaque cerebral cortex has many visual areas that can be distinguished on the basis of connectivity, architectonics and receptive field organization. In a well-known review paper, Felleman and Van Essen¹ arranged these areas in a rough hierarchy, with area V1 (also called the striate cortex or area 17) at the bottom. It is a unique feature of the cerebral cortex that it consists of regions of quite similar structure than can be connected in series. This is in marked contrast to other major brain systems, such as the thalamus, the basal ganglia and the cerebellum, which deal with many different parallel inputs but do not form repeating serial sets of connections. The different regions of the thalamus, for example, are not directly connected to each other in series.

In the mammalian visual system, a cortical area higher in this serial hierarchy can form representations of complex visual objects that are built up from less complex visual features at lower stages. The ability to add additional areas to such a hierarchy, while still allowing the system to learn, is probably the secret of the success of the cerebral cortex in evolution.

Felleman and Van Essen¹ constructed their hierarchy by using a set of rules (Fig. 1) based on an earlier suggestion². A connection between two cortical areas is considered as ascending in the

hierarchy if it terminates mainly in cortical layer 4, especially if it originates in the supragranular cortical layers (layers 2 and 3). It is considered as descending in the hierarchy if it avoids layer 4, and terminates strongly in layer 1 and possibly in layer 6 as well. These are their strong rules.

They also considered what they called lateral connections, which connected cortical areas at the same level in the hierarchy, for instance between areas MT and V4. The connections could originate in all layers that project out of the cortical area (that is, from all layers except layers 1 and 4) and could terminate in all layers. This is their weak rule. Here we will consider only their strong rules.

Felleman and Van Essen¹ placed the (mainly) visual areas on a hierarchy having ten distinct levels, but they admitted that there were irregularities. In their words, these irregularities "raise the issue of whether the cortex is inherently only a 'quasi-hierarchical' structure that contains a significant number (perhaps 10%) of bona fide irregularities and exceptions to any set of criteria that can be devised. Alternatively, the visual cortex might contain an essentially perfect anatomical hierarchy that has been imperfectly studied using inherently 'noisy' methods of anatomical analysis."

The idea of a unique hierarchy has been criticized by Hilgetag,

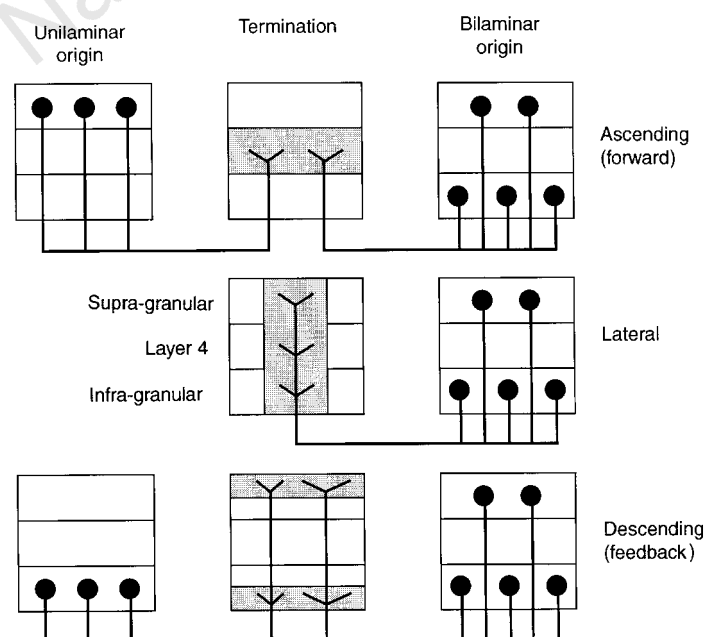


Figure 1 Rules used by Felleman and Van Essen¹ for deciding whether a projection between two cortical areas in the macaque monkey is ascending, descending or lateral (based on Fig. 3 of ref. 1).

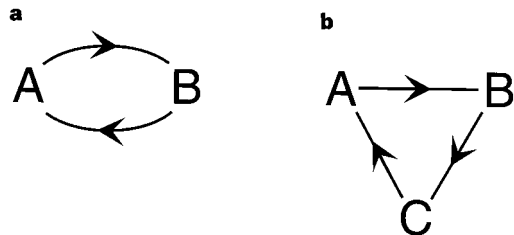


Figure 2 Forbidden cortical loops. A, B and C represent distinct cortical areas. Arrows represent strong driving connections into cortical layer 4. **a**, Binary loop; **b**, tertiary loop. The examples shown are not found in the macaque visual cortex, at least in the lower parts of the visual hierarchy.

O'Neill and Young³, who have used an evolutionary optimization algorithm to find hierarchies that have the fewest departures from a perfect hierarchy. They showed that many hierarchies were possible, each with a number of violations of the rules. They noted that if 36 of the less-reliable constraints were excluded, then there were only two violations. They concluded, "The primate cortical visual system is therefore surprisingly strictly hierarchical, but it is nonetheless not possible to determine the exact hierarchy."

The no-strong-loops hypothesis

We address the question: are there any constraints on the type of connections that can be used to form a hierarchical system? Our suggestion is that the system must avoid strong connections that connect areas in a closed loop.

What do we mean by 'strong connections'? In quantitative terms, this is, at present, a vague idea. In qualitative terms, we would like to believe that an appreciable connection into layer 4 is a driving connection (which we consider to be strong), and a connection to layer 1—mainly to the apical dendrites of pyramidal cells—is a modulating connection (which we consider to be weak). By driving inputs, we mean one or more inputs that, by themselves, can make the relevant neurons fire strongly. By modulating inputs, we mean one or more inputs that, by themselves, cannot make the relevant neurons fire strongly, but can modify the firing produced by the driving inputs. It can be argued that the driving inputs contribute mainly to the classical receptive field of a neuron, and the modulating inputs largely to the non-classical receptive field.

We shall call the first type of connection a D_{cc} connection and the second an M_{cc} connection (the subscript cc means cortex-to-cortex). Our cortical hypothesis is that if we consider only D_{cc} connections between visual cortical areas, they will never form a directed loop, either a binary loop (Fig. 2a) or a tertiary loop (Fig. 2b), nor any higher-order loop. Notice that we are talking about connections between cortical areas, not connections between individual neurons.

Close inspection of the data of Felleman and Van Essen¹ shows that this rule is always obeyed. As far as we know, there does not appear to be any more recent experimental evidence that contradicts it, at least for the earlier parts of the visual hierarchy, although the evidence for the later parts is very skimpy.

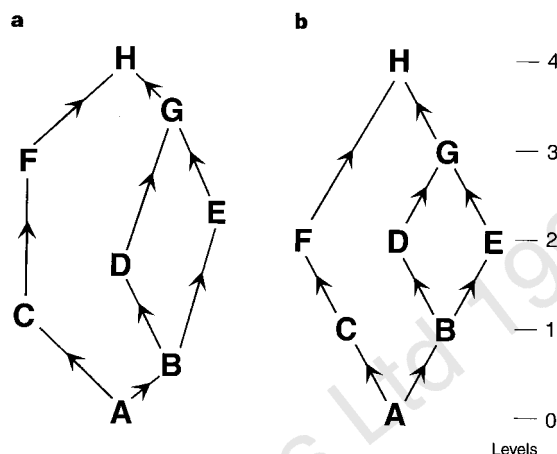


Figure 3 An acyclic digraph with eight nodes (A to H), drawn so that all the arrows point upwards. In **a**, the nodes are all drawn at different levels. Many other arrangements of levels are possible. In **b**, these nodes have been drawn using the minimum possible number of levels (five). In this case, there are two other possible ways of doing this. The one shown exhibits the unique numbering of levels used by mathematicians. (Notice that it would make no difference to the levels if there were an additional arrow from A to D.)

Digraphs

A hierarchy without directed loops is not necessarily unique. This follows from the mathematical theory of directed graphs, or digraphs as they are often called⁴. A digraph consists of a series of nodes (in our case, cortical areas) connected by lines, each line having an arrow on it to show its direction. If a digraph has a directed loop, then it is not possible for the nodes to form a hierarchy of levels. If it has no directed loop, the nodes can form a hierarchy, but it may not be a unique one.

An example may make this clearer. Consider the digraph in Fig. 3a: it is easy to see that, for example, nodes B, D, E and G could all be drawn at levels above F (but below H), or that they could all be drawn below C (but above A). Many other arrangements of the exact levels of the nodes are possible, but each one is a possible hierarchy and none of them has directed loops. Even if one adds the constraint that the number of levels should be kept to a minimum⁵, there are in this case three possible hierarchies, each with five levels (one is shown in Fig. 3b). However, there is a unique way of numbering the levels of an acyclic digraph: each node is assigned an integer equal to the number of links in the longest path from the lowest level to it (Fig. 3b).

We will postpone a more general discussion of possible reasons for our hypothesis until we have considered a possible extension of it to connections between the cortex and the thalamus.

The visual thalamus

The dorsal visual thalamus in primates consists mainly of the well-known lateral geniculate nucleus (LGN) and the less-studied but much larger pulvinar. The visual pulvinar has at least three major parts: the inferior, the lateral and the medial pulvinar. Unfortunately, these divisions are not completely clear-cut because the evidence is very patchy.

This is partly because the divisions are sometimes based on architectonics, sometimes on their connections, and sometimes on physiological properties. Moreover, each of these major regions probably contains subregions. For example, a chemoarchitectonic (staining) description of the macaque inferior pulvinar⁶ (as defined by the authors) suggested that it had five possible subdivisions. It is likely that half a dozen or more separate maps of the visual environment exist in the various pulvinar nuclei⁷.

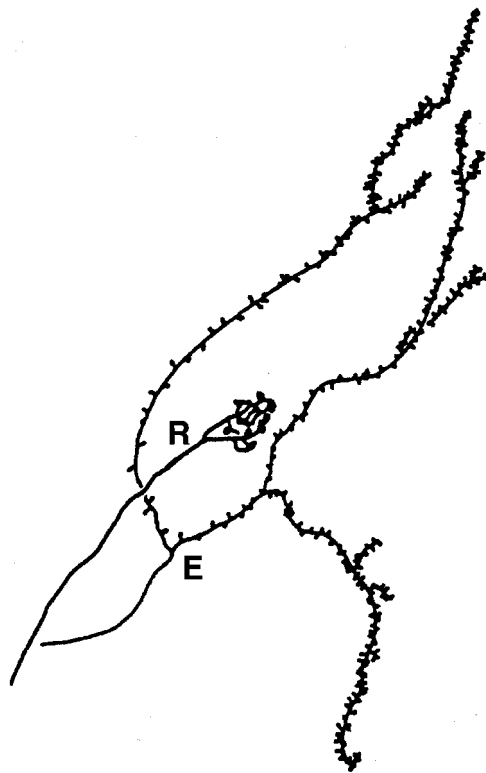


Figure 4 Termination of a typical R (round)-type corticopulvinar axon and an E (extended) corticopulvinar axonal termination. The diameter of the terminal field of the R-type axon is $\sim 100 \mu\text{m}$. Based on data from the macaque monkey (taken from part of Fig. 24 of Rockland¹²).

There is good evidence that the asymmetric synapses formed by afferent inputs to the thalamus are of two main types^{8–15}. The axon terminals of one type are invariably concentrated in sharply delineated arbors; Rockland¹² calls these R (round)-type terminations. The other type are densely arrayed along axons of slightly smaller calibre, which Rockland calls E (extended). The two types are shown in Fig. 4.

R-type terminals are beaded and characteristically large ($3 \mu\text{m}$ in diameter), although variable in size and actual shape. They conform to the classical type-2 endings, as described in specific thalamic nuclei¹⁶. The associated axonal terminations are concentrated in sharply delimited, round arbors and carry of the order of 100 terminals, that typically end on proximal dendrites.

E-type axons have stalked or spinous terminations of classic type-1 corticothalamic endings¹⁶. Their axonal terminal fields are elongated and quite extended ($1–3 \text{ mm}$) and carry between 500 and 1,000 E terminals that typically end on distal dendrites.

In the LGN, the driving input from the retina is provided by R-type axon terminals, with type-2 synapses; the input back from cortical area V1 has E-type axon terminals, with type-1 synapses. We shall tentatively assume that the latter is a modulating input, in spite of the fact that there are many more E-type than R-type axons. In part, this is based on the fact that eliminating or transiently inactivating primary visual cortex has only weak, secondary effects on the receptive field properties of geniculate cells^{17–19}.

More recent evidence^{11,12,14} shows that the same two types of axon (and synapse) are also found in the macaque pulvinar. Thus we shall generalize our assumptions by postulating that the R-type is a driving connection (which we shall call d_{ct} , where the subscript means 'cortex to thalamus'), and that the E-type (though more numerous) is a modulating connection (which we shall call m_{ct}). The properties of the two types of corticothalamic connections are

summarized in Table 1. It is highly likely that all these connections are excitatory, using glutamate as neurotransmitter.

The E-type cortical axons probably derive from medium to small pyramidal cells in the lower cortical layers. The evidence available suggests that they are located in layer 6, and as a rule always have collaterals in the thalamic reticular nucleus. There is suggestive evidence^{12,13,20} that cortical R-type axons originate from pyramidal cells in cortical layer 5. In the case of rat V1, some of these cells project to the tectum and the pontine nucleus, whereas for V1 in the macaque, some of these layer-5 cells branch and project to both superior colliculus and the pulvinar²¹.

Thalamic projections to the cortex

Projections from the thalamus to the cortex also fall into two classes. The first type goes mainly into layer 4 or lower layer 3, with a minority also contacting processes in layer 6. The projection cells in the magno- and parvocellular laminae of the monkey LGN are prominent examples of such a connection that can very reliably drive cortical cells, despite their small number of synapses²². Note that a large number of synaptic inputs are not needed to drive a cell reliably. As few as 2.8% of all excitatory synapses on a layer 4C α spiny stellate cell originate from magnocellular cells in LGN²². To us, this is a prime example of a driving connection, so we will call it d_{tc} (where the subscript means 'thalamus to cortex'). The other type projects to layer 1, but not exclusively. We postulate that this is a modulating connection and call it m_{tc} . An example of this is shown by cells in the interlaminar zones of the LGN that project into the superficial layers of V1²³.

We can now state our extended hypothesis: in the primate visual system, if we consider only what we have called driving connections (that is, corticocortical connections of type D_{cc} , corticothalamic connections of type d_{ct} , and thalamocortical connections of type d_{tc}), then these connections never form a strong directed loop.

This extended hypothesis is supported by Jones' rule, put forward by Jones in his monograph on the thalamus²⁴. He suggested that connections between the cortex and the thalamus usually fall into two distinct classes, as shown in Fig. 5a. (1) If a cortical area projects to a thalamic region from cortical layer 6, then if there is a reverse projection, it goes mainly into layer 4 or lower layer 3; (2) if a cortical area projects to a thalamic region from cortical layer 5, then if there is a reverse projection it avoids layer 4 and often goes mainly to cortical layer 1. These thalamocortical projections are usually much more diffuse than the layer 4 projection.

It is not clear at the moment whether Jones' rule is always obeyed. Jones' rule is certainly obeyed in the case of V1. Here, the cortico-geniculate pathway originates in layer 6, whereas most geniculate axons terminate in layer 4 (ref. 15). The projection from V1 to the pulvinar originates in layer 5; the backprojections from the pulvinar to V1 appear to end superficially in layers 1 to 3, rather than in layer 4.

There is evidence that some of the projections from parts of the pulvinar to extrastriate cortex end in layers 3 and 4, and others mainly in layer 6, avoiding the middle layers²⁴. The precise details of the connections between the many visual cortical areas and the various subdivisions of the pulvinar are not known, except that the corticopulvinar pathways always originate in an infragranular layer^{12,25}.

What might Jones' rule mean in our terminology? Let us assume (as seems likely) that projections from cortical layer 6 are formed by

Table 1 The two main types of corticothalamic connections*

Axons	Synapses	Location on dendrites	Character
R type: few	Type 2, beaded	Proximal	Driving (?)
E type: many	Type 1, stalked	Distal	Modulating (?)

* See Fig. 4.

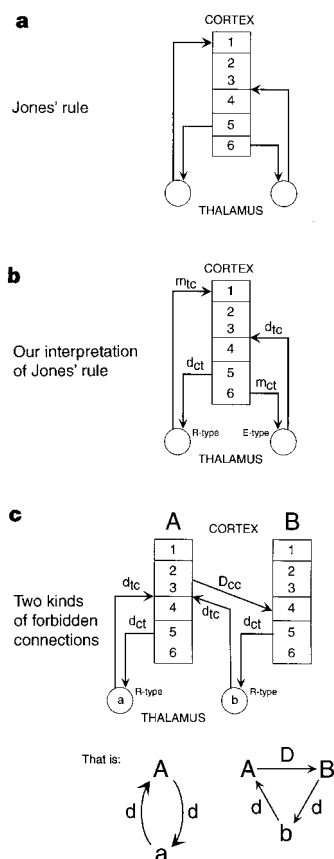


Figure 5 Permissible and forbidden connections between thalamus and cortex. **a**, Jones²⁴ described the two types of connections between thalamus and cortex he had observed, based on their layer of origin and termination. Rule 1 is illustrated on the right; rule 2 on the left. **b**, Our interpretation of these rules in terms of driving connections (d_{ic} into cortical layer 4 or d_{ct} from the lower layers and having R-type synapses) and modulating connections (m_{ic} into layer 1 or m_{ct} from layer 6 and having E-type synapses). **c**, Our no-strong-loops hypothesis rules out these types of strong connection between the thalamus and cortex. (In this pair of diagrams, the modulating pathways have been omitted to avoid confusion.)

E-type axons—which we called an m_{ct} connection—and projections from layer 5 by R-type axons, which we called a d_{ct} connection. Figure 5b shows our interpretation of Jones' rule. Jones' rule does not deal with the connections shown on the left of Fig. 5c, which is forbidden by our hypothesis, because it has two driving connections in a loop. In short, Jones' rule supports our hypothesis.

However, our hypothesis also extends to tertiary and higher interactions, where it states we should never find connections as those shown on the right-hand part of Fig. 5c. To test our extended hypothesis properly, it will be necessary to have a clean division of the pulvinar into distinct regions, and more experimental data on exactly how these various regions connect to the many cortical areas in the visual cortex.

Strictly speaking, our hypothesis applies to the type of axon (R or E), whereas Jones' hypothesis considers the cortical layer of origin. Although R-type cortical axons probably arise from layer 5 in most cases, there may be exceptions.

In formal terms, then, what we predict is that if a certain thalamic region, x , projects strongly to the middle layers of a certain cortical area, Z , then any projection x receives from R axons must come from a cortical area Y , which is lower in the cortical hierarchy than Z . That is, Z should not project (directly, or through cortical intermediaries) strongly into the middle layers of Y .

It might be considered that with so many connections between the cortex and the pulvinar, it would be difficult to avoid forming at least a few strong directed loops: this need not be the case. Noting that the postulated driving connection from layer 5 of V1 to the pulvinar is in an upward direction in the visual hierarchy, we can avoid directed loops if none of the strong connections projects downwards in the hierarchy. A simple example of such a set of permissible driving connections is shown in Fig. 6.

This hypothetical example hints that the strong connections are all of the feedforward type, and that the visual cortex is basically a feedforward system that is modulated by the various feedback connections. (This is not to say that such modulation may not be very important for many of its functions.) Whether this overall view of the macaque visual system is correct remains to be discovered. Notice that some of these postulated strong pathways go through the pulvinar, as has already been stressed¹⁴. This aspect of pulvinar connections has often been ignored.

Unfortunately, it is unclear what should be considered a single region of the pulvinar. It may be a fairly large volume defined, for example, by its staining properties⁶. However, from our point of view it might have a smaller volume, because the extent of axonal collaterals of thalamocortical cells within the thalamus appears to be minor, in contrast to cortex where intracortical connectivity can be quite extensive. Thus, to some extent, a small volume of the pulvinar might act largely independently of neighbouring volumes, and so might be considered a 'region' from the point of view of our hypothesis.

As there are likely to be fewer large regions of the pulvinar than cortical areas to which it connects, it will be of interest to discover exactly which subset of cortical areas receives a d_{ic} connection from a particular pulvinar region. It is possible that all of these cortical areas can be considered to occur at one particular level in the visual hierarchy.

Theory

Assuming that further experimental evidence confirms our no-strong-loops hypothesis, what could be the reason for it? The obvious explanation is that a strong excitatory loop would throw the cortex into uncontrolled oscillations, as in epilepsy. It could be contended that there is enough local inhibition in the system to choke off such oscillations, but the fact that epilepsy does occur suggests there are limits to the control that inhibition can have over nascent oscillations. We thus think it likely that the reason no strong loops are observed (if this is true) is that the system could not constrain their effects, possibly because neurons are usually kept in a jittery state near the edge of excitation in order to make them respond more promptly, so that the amount of inhibition in the system is limited. The obvious test would be to construct a strong loop in some way, but this is beyond our present experimental capabilities.

It is well known to neural-net modellers (H. White, personal communication) that although neural nets can be constructed with feedback connections that form loops, they do not work satisfactorily if the excitatory feedback is too strong (for several special cases, see refs 26, 27). This was also true for early radio sets that emitted howls (caused by positive feedback) if the volume control was turned up too high. However, it is almost impossible at the moment to use detailed neural network models to prove our hypothesis, because too little is known about many of the connections and parameters involved.

Possible generalizations

Our hypothesis is sufficiently plausible that it could be extended to other parts of the nervous system. We have considered only the visual system of primates, but it could be extended to all parts of the cerebral cortex, and to all parts of the thalamus, including the intralaminar nuclei, and to other mammals and other vertebrates. It

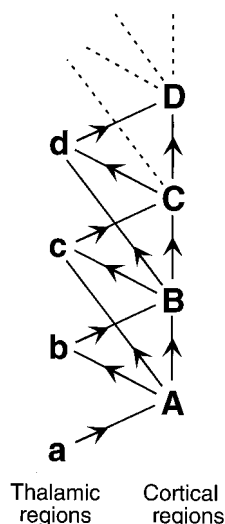


Figure 6 A hypothetical set of connections lacking strong directed loops. A, B, C and D represent cortical areas; a, b, c and d represent thalamic regions. Only strong (driving) connections are shown. Because no arrow points downwards in the hierarchy, no strong directed loops can occur. It is easy to add further strong connections, provided they all point upwards. Note that this diagram suggests that each thalamic region can be assigned a unique level.

might be instructive to consider to what extent connections of the superior colliculus, the claustrum, the cerebellum and the basal ganglia into or out of the cortex and thalamus can be classified as driving or modulating.

Does our no-strong-loops hypothesis rule out the existence of a significant number of directed loops between two individual neurons (as opposed to areas)? An example of such a two-neuron loop would be a thalamocortical projection neuron making a driving connection in layer 4 onto the apical dendrite of a layer-5 cell that drives the thalamic cell. We leave all these topics for further consideration.

Discussion

Our division of the connections into 'driving' and 'modulating' connections may well be too simple. In the first place, although it is probably that a 'driving' connection does excite the recipient neurons to fire, it is less certain that a 'modulating' connection can only modulate and not, under some circumstances, fire the recipient neurons by itself. This is why we called them D/d and M/m, to indicate that our idea of their exact function is only rather tentative. (See experiments on the anaesthetized squirrel monkey²⁸ and on the anaesthetized cat²⁹, and a critique of these experiments³⁰.) However, under natural circumstances, we think that the D/d connections are likely to produce a stronger excitation than the M/m ones. Also, the timescale of their effect might differ, with the M/m pathways acting as a 'bias' over a much longer timescale than the D/d pathways. The exact nature of the receptors involved might also be different.

Our general point is that all the different types of connection should not be regarded as having the same character and the same 'strength'. Although it is not true that "everything is connected to everything else", the connections between cortical areas are very extensive (see discussion in ref. 1). It is therefore important to try to classify the connections into different types, in the hope that by doing so the behaviour of the system can be comprehended more easily. Connections may differ in many important ways apart from their layer of origin and destination: for example, how widely their

axons arborize, the type and dendritic location of the synapses they form, the receptors involved, and so on.

An analogy may help here. In organic chemistry, the atoms making up one molecule are bonded together and also, less strongly, to the atoms in neighbouring molecules. Although the details of the different bonds vary from atom to atom, it is useful to distinguish strong bonds, such as homopolar and coulombic bonds, from weak bonds, such as hydrogen bonds and those due to van der Waals forces. Moreover, each of these types of bonds has its own character. It would be impossible to understand the structure of, say, a protein molecule if all these 'bonds' were considered to be on an equal footing.

In the same way, one should attempt to classify the interactions between neuronal groups (such as those in a single cortical area) into different types, in order to obtain a general picture of what is going on. Our classification above should be regarded as a tentative step in that direction.

Are there likely to be significant exceptions to our no-strong-loops hypothesis? It is possible that most directed loops are weak loops but that there may be some loops that are of intermediate strength. These might allow sustained activity of finite duration. Not surprisingly, we have wondered whether such intermediate-strength loops might form the neural correlate of consciousness; in particular, we would draw attention to some of the loops between the frontal cortex and the higher visual cortical areas³¹. There is already a hint that the laminar connections there are not what might have been expected. A paper by Ungerleider and colleagues³² stated, "... it may be that the rules that have been used for establishing hierarchical relationships within both the visual and somatosensory systems¹ do not extend, in any simple way, to connections with frontal lobe areas."

Conclusions

Our hypothesis is relatively easy to test experimentally, at least in its neuroanatomical form. It is based on certain neuroanatomical connections that have already been established and predicts the general nature of other connections that have not yet been established definitively. It seems very likely that it is true for the corticocortical connections in at least the lower levels of the macaque visual hierarchy. More experimental data are needed to test it for the higher levels of the visual hierarchy (and especially the connections to frontal cortical areas) and for the many connections between the pulvinar and the visual cortex.

Although neuroanatomy can by itself confirm our prediction of the absence of certain connections, it cannot confirm our explanation for these absences unless there is good neurophysiological evidence that what we have called driving connections are in some significant sense stronger than the connections we have termed modulating. 'Strength' implies that a strong connection fires those neurons more effectively than a weak one. It is possible that the reason for the strength of connections to cortical layer 4 is the amplification produced by local lateral connections³³. The evidence suggests that the strength of a connection cannot be gauged by merely counting the number of axons involved. In the long run, we need to know not only the 'strength' but something of the general character of the effects produced by each type of connection; for instance, whether they originate from neurons that fire in bursts³⁴.

If our cortical hypothesis is true, it might explain why the hierarchy in the visual cortical areas may be only a rough one, and would call a truce on discussions as to which version of the hierarchy is the best. If our extended hypothesis, which includes the thalamic connections, turns out to be true, it may suggest an unambiguous hierarchy for at least most of the visual cortical areas. Our hypothesis is, however, a negative one: it predicts which types of connections will not be found. It says very little about the complex processing, involving all the types of connections, that allows an animal to see and to act on what it sees. It does

not explain the function of modulatory m_{ct} or m_{tc} connections or of corticocortical feedback connections. To study this, we need much further work and probably new ideas; our hypothesis merely attempts to clear the ground for such endeavours. □

Received 3 July; accepted 10 October 1997.

1. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
2. Rockland, K. S. & Pandya, D. N. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res.* **179**, 3–20 (1979).
3. Hilgetag, C.-C., O'Neill, M. A. & Young, M. P. Indeterminate organization of the visual system. *Science* **271**, 776–777 (1996).
4. Robinson, D. F. & Foulds, L. R. *Digraphs: Theory and Techniques* (Gordon and Breach, New York, 1980).
5. Van Essen, D. C. & Felleman, D. J. On hierarchies (response to Hilgetag *et al.*). *Science* **271**, 777 (1996).
6. Gutierrez, C., Yaun, A. & Cusick, C. G. Neurochemical subdivisions of the inferior pulvinar in macaque monkeys. *J. Comp. Neurol.* **363**, 545–562 (1995).
7. Robinson, D. L. & Cowie, R. J. in *Thalamus, 2: Experimental and Clinical Studies* (eds Steriade, M., Jones, E. G. & McCormick, D. A.) (Elsevier, Amsterdam, in the press).
8. Mathers, L. H. Ultrastructure of the pulvinar of the squirrel monkey. *J. Comp. Neurol.* **146**, 15–42 (1972).
9. Mathers, L. H. The synaptic organization of the cortical projection to the pulvinar of the squirrel monkey. *J. Comp. Neurol.* **146**, 43–60 (1972).
10. Ojima, H. Terminal morphology and distribution of corticothalamic fibers originating from layers 5 and 6 of cat primary auditory cortex. *Cereb. Cortex* **6**, 646–663 (1994).
11. Rockland, K. S. Further evidence for two types of corticopulvinar neurons. *NeuroReport* **5**, 1865–1868 (1994).
12. Rockland, K. S. Two types of corticopulvinar terminations: round (type 2) and elongate (type 1). *J. Comp. Neurol.* **368**, 57–87 (1996).
13. Bourassa, J., Deschênes, M. Corticothalamic projections from the primary visual cortex in rats: a single fiber study using biocytin as an anterograde tracer. *Neuroscience* **66**, 253–263 (1995).
14. Guillery, R. W. Anatomical evidence concerning the role of the thalamus in corticocortical communication: a brief review. *J. Anat.* **187**, 583–592 (1995).
15. Sherman, S. M. & Guillery, R. W. Functional organization of thalamocortical relays. *J. Neurophys.* **76**, 1367–1395 (1996).
16. Guillery, R. W. Patterns of synaptic interconnections in the dorsal lateral geniculate nucleus of the cat and monkey: a brief review. *Vision Res.* **3**, 211–227 (1971).
17. Kalil, R. E. & Chase, R. Corticofugal influence on activity of lateral geniculate neurons in the cat. *J. Neurophysiol.* **33**, 459–474 (1970).
18. Baker, F. H. & Malpeli, J. G. Effects of cryogenic blockade of visual cortex on the responses of lateral geniculate neurons in the monkey. *Exp. Brain Res.* **29**, 433–444 (1977).
19. Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* **369**, 479–482 (1994).
20. Callaway, E. M. Local circuits in primary visual cortex of the macaque monkey. *Annu. Rev. Neurosci.* (in the press).
21. Rockland, K. S. Convergence and branching patterns of round, type 2 corticopulvinar axons. *J. Comp. Neurol.* (in the press).
22. Peters, A., Payne, B. R. & Budd, J. A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb. Cortex* **4**, 215–229 (1994).
23. Hendry, S. H. C. & Yoshioka, T. A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. *Science* **264**, 575–577 (1994).
24. Jones, E. G. *The Thalamus* (Plenum, New York, 1985).
25. Levitt, J. B., Yoshioka, T. & Lund, J. S. Connections between the pulvinar complex and cytochrome oxidase-defined compartments in visual area V2 of macaque monkey. *Exp. Brain Res.* **104**, 419–430 (1995).
26. Hirsch, M. W. Convergent activation dynamics in continuous time networks. *Neural Networks* **2**, 331–349 (1989).
27. Kuan, C.-M., Hornik, K. & White, H. A convergence result for learning in recurrent neural networks. *Neural Comp.* **6**, 420–440 (1994).
28. Sandell, J. H. & Schiller, P. H. Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J. Neurophys.* **48**, 38–48 (1982).
29. Mignard, M. & Malpeli, J. G. Paths of information flow through visual cortex. *Science* **251**, 1249–1251 (1991).
30. Salin, P.-A. & Bullier, J. Corticocortical connections in the visual system: structure and function. *Physiol. Rev.* **75**, 107–154 (1995).
31. Crick, F. & Koch, C. Are we aware of neural activity in primary visual cortex? *Nature* **375**, 121–123 (1995).
32. Webster, M. J., Bachevalier, J. & Ungerleider, L. G. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex* **5**, 470–483 (1994).
33. Douglas, R., Koch, C., Mahowald, M., Martin, K. & Suarez, H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
34. Lisman, J. E. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* **20**, 38–43 (1997).

Acknowledgements. E.C. is supported by the J. W. Kieckhefer Foundation. C.K. is supported by the National Institute of Mental Health, the Office of Naval Research and the National Science Foundation. Especial thanks to Jan Kaas and Kathy Rockland.

Correspondence and requests for materials should be addressed to E.C.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

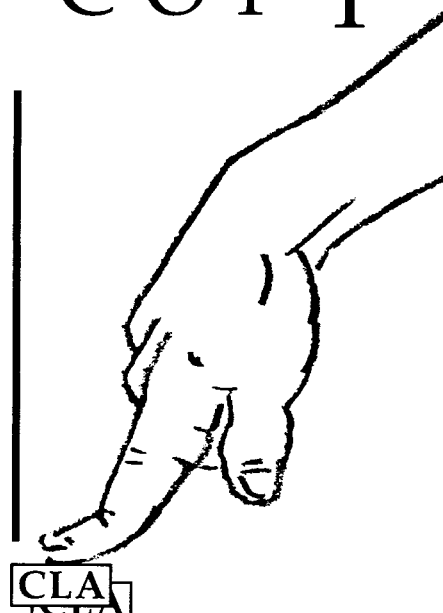
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution

Sylvie Doublé, Stanley Tabor, Alexander M. Long, Charles C. Richardson & Tom Ellenberger

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

DNA polymerases change their specificity for nucleotide substrates with each catalytic cycle, while achieving error frequencies in the range of 10^{-5} to 10^{-6} . Here we present a 2.2 Å crystal structure of the replicative DNA polymerase from bacteriophage T7 complexed with a primer-template and a nucleoside triphosphate in the polymerase active site. The structure illustrates how nucleotides are selected in a template-directed manner, and provides a structural basis for a metal-assisted mechanism of phosphoryl transfer by a large group of related polymerases.

Genomic DNA is replicated by polymerases working in concert with accessory proteins that unwind the DNA duplex, increase the processivity of polymerization, prime the lagging strand synthesis, and coordinate the synthesis of leading and lagging strands. The replication system of bacteriophage T7 is an attractive model for studying the dynamics of a replication fork, in part because only a few proteins are involved. The 80K DNA polymerase encoded by gene 5 of the phage is a distributive enzyme that dissociates from a primer-template after incorporating only 1–15 nucleotides¹. During phage growth, the polymerase annexes *Escherichia coli* thioredoxin as its processivity factor² forming a tight one-to-one complex ($K_d \sim 5$ nM) that polymerizes thousands of nucleotides before dissociating from the primer-template¹. Additional interactions between the polymerase, the hexameric T7 primase-helicase, and the T7 single-stranded DNA binding protein coordinate the synthesis of leading and lagging DNA strands at the replication fork³. As a first step towards defining the three-dimensional structure of the T7 replisome, we have determined a crystal structure of T7 DNA polymerase complexed to thioredoxin, a primer template, and a nucleoside triphosphate. The fold of the T7 DNA polymerase resembles that of other members of the *E. coli* DNA polymerase I (Pol I) family^{4,5}, but the subdomains surrounding the polymerase active site are configured in an unprecedented closed conformation that secures the 3' terminus of the primer and positions the nucleoside triphosphate in the active site. This structure provides a number of mechanistic insights into the chemistry and fidelity of DNA synthesis.

Structure determination of the complex

T7 DNA polymerase forms a stable complex with a primer-template in which the primer strand is terminated by 2',3'-dideoxynucleotide, provided that the next dNTP specified by the template is present (S.T. and C.C.R., manuscript in preparation). For crystallization, T7 DNA polymerase was mixed with a 21 base-pair duplex having a single-stranded four-base extension at the 5'-end of the template strand, in buffer containing 2',3'-dideoxyadenosine triphosphate (ddATP) and 2',3'-dideoxyguanosine triphosphate (ddTGP) (see Methods). After incorporating ddAMP into the primer strand, the polymerase remained bound to the primer-template and ddGTP in a complex having a half-life of 1 h. An analogous strategy has been used to crystallize mammalian DNA polymerase- β complexed to a primer-template and a nucleoside triphosphate⁶. To prevent degradation of the primer-template by the 3'–5' exonuclease activity of T7 DNA polymerase, a number of exonuclease-deficient polymerases were screened for crystallization.

The best crystals were obtained with a polymerase lacking six residues (118–123) in the exonuclease domain. The exonuclease activity of this altered enzyme is decreased by several orders of magnitude, yet it has normal polymerase activity and it supports the growth of T7 phage⁷. The structure of the polymerase complex was determined at 2.2 Å resolution using multiple wavelength anomalous diffraction from crystals of selenomethionine-substituted protein (Table 1 and Methods). The current *R*-factor is 0.240 for the model refined against all data (20 Å–2.2 Å resolution; $R_{\text{free}} = 0.279$).

T7 DNA polymerase has an architecture that is characteristic of other polymerases in the Pol I family, with an amino-terminal 3'–5' proofreading exonuclease domain and a carboxy-terminal polymerase domain. The polymerase domain has a shape reminiscent of a right hand⁸ in which the palm, fingers and thumb form a DNA-binding groove that leads to the polymerase active site (Fig. 1). The orientation of the primer template places the primer 3'-end against the ddGTP at the confluence of the fingers and palm. The palm forms the base of the polymerase active site, presenting three strictly conserved^{4,5} and functionally important^{9,10} acidic residues. Two metals are held in the active site by two of these conserved carboxylates, and these metals in turn ligate the phosphates of the bound ddGTP (Fig. 2). The fingers wall off this end of the DNA-binding groove, making numerous interactions with the bound nucleotide (Figs 1 and 2). Crystal structures of other Pol I family members reveal a similar fold of the polymerase domain, including that of the Klenow fragment of *E. coli* Pol I^{11,12}, the analogous fragments of *Thermus aquaticus* DNA polymerase (Taq DNA polymerase)¹³ and *Bacillus stearothermophilus* DNA polymerase¹⁴, and intact Taq polymerase¹⁵. This conservation of structure is most pronounced in the palm, where β -strands 9, 12 and 13 of T7 DNA polymerase superimpose very well on the analogous strands of Klenow fragment^{11,12} (r.m.s. deviation = 0.84–0.88 Å), and Taq polymerase^{13,15,16} (r.m.s. deviation = 0.78–0.80 Å). The 3'–5' exonuclease domain lies at the end of the DNA-binding groove away from the fingers, in an orientation similar to that in the Klenow fragment^{11,12}. In T7 DNA polymerase, a 71-residue insertion in the tip of the thumb forms an extended loop that binds to thioredoxin and is poised over the duplex DNA exiting the polymerase (Fig. 1a).

DNA distortions

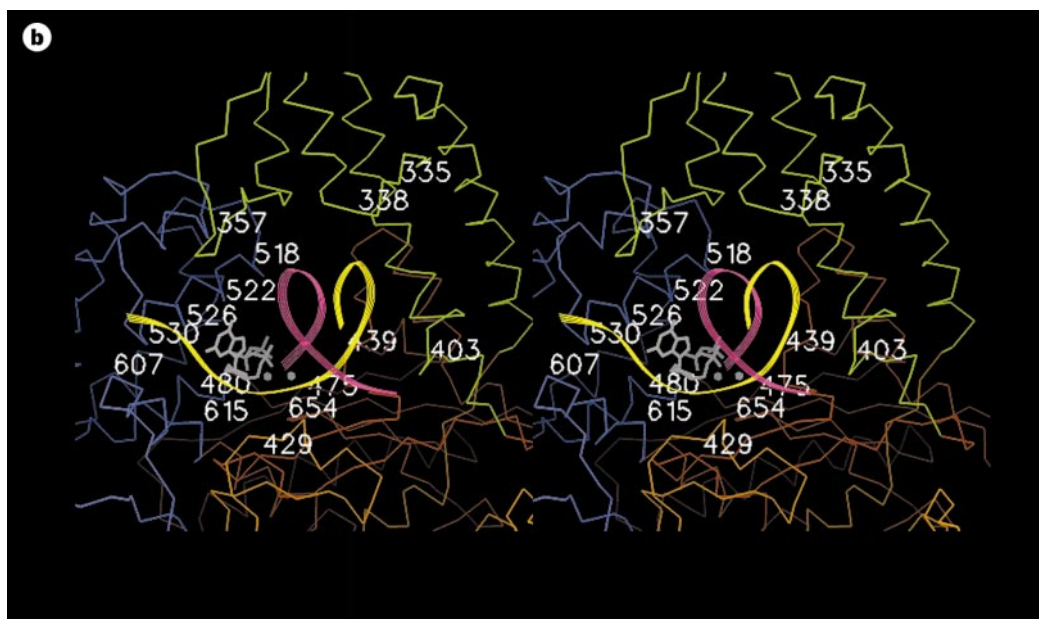
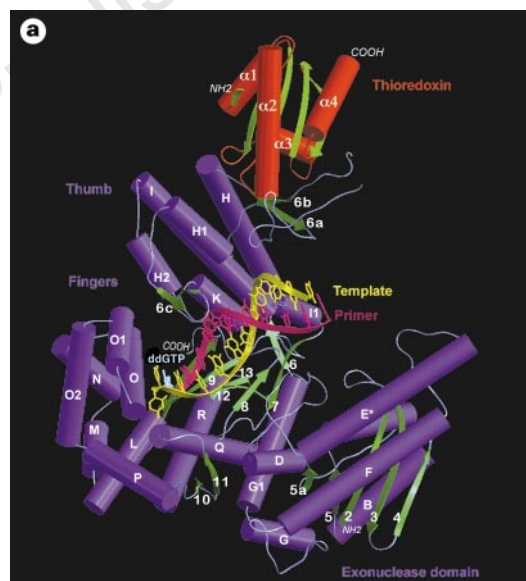
The DNA bound by the polymerase is predominantly in the B-form. However, two base pairs at the primer 3'-end are A-form, with a decreased helical twist and a widened minor groove. A similar transition from B- to A-form has been described for DNA com-

plexed with Taq DNA polymerase¹⁶ and DNA polymerase- β (ref. 6). In contrast to these latter two structures, the DNA in the T7 DNA polymerase complex is bent into an S-shape by extensive interactions with the thumb and fingers (Fig. 1b). A curve in the DNA next to the active site compresses the major groove, and the thumb imposes a second bend in the DNA that completes the S-curve. Eleven base pairs of the duplex extend outside the polymerase and are disordered. The single-stranded 5'-end of the template does not pass through the crevice between the thumb and fingers, but instead lies on the surface of the fingers (Fig. 1). The cytidine preceding the template base is flipped out of the active site and it stacks against His607 (Fig. 2). The two 5'-terminal cytidines of the single-stranded template are disordered. Thus, the template strand enters the polymerase active site from the side away from the thumb, and a sharp turn exposes the template base for interaction with the incoming nucleotide (Fig. 1a). The crystal structure of polymerase- β complexed to a nucleoside triphosphate and a duplex DNA with a single nucleotide gap also shows a sharp turn at the template base¹⁷.

The primer-template is held in position by direct and water-

mediated interactions of the fingers, palm and thumb, primarily to the phosphodiester backbone (Fig. 2). The few contacts made to the bases appear to be at positions critical for ensuring high fidelity, and they will be discussed below. The fingers contact both the primer and template strands adjacent to the active site. The primer 3'-end is anchored by the fingers and palm and its 5'-end is held by the thumb. The thumb pushes against the minor groove of the DNA four to five base pairs from the active site, with numerous interactions made by helices H1 and H2 and nearby loops at the tip of the thumb, and helix I at the base of the thumb. Asn 335 and Ser 338 are part of a conserved sequence motif⁴ in the loop preceding helix H1 that contacts the template strand. The segment connecting helices H1 and H2 is well ordered and it includes two short antiparallel β -strands that drape over the primer, which is also contacted by residues near the N terminus of helix H2 (Figs 1 and 2). Deletion of 24 residues from this region at the tip of the Klenow fragment thumb decreases DNA-binding affinity and promotes frameshift mutations¹⁸. A four-residue loop located between helices I and II at the base of the thumb contacts the template strand. Asp 403 and Lys 404 within this loop make direct and water-mediated contacts to

Figure 1 Topology and overall fold of T7 DNA polymerase. **a**, Schematic representation in which helices are represented by cylinders, and β -strands by arrows. Secondary structure elements were defined by PROCHECK⁴⁷. The labelling of the structural elements follows that of Klenow fragment^{11,12}, in which helices are represented by letters and β -strands by numbers. Residues 294 to 317 in the thioredoxin-binding domain are disordered in the crystal structure, as are residues 576 to 586 between helix P and β -strand 10 of the fingers. The secondary structure assignments of the polymerase are as follows. Exonuclease domain: 2 (1-9), 3(17-26), 4(30-34), B(38-52), 5(55-59), C(64-77), 5a(87-90), D(91-99), E'(126-150), F(164-187). Polymerase domain: G(202-210), G1(211-224), 6(231-234), H(235-261), 6a(264-268), 6b(327-332), H1(339-348), 6c(355-357), 6d(361-363), H2(365-372), I(376-400), I1(405-410), 7(414-417), 8(430-433), K(447-454), 9(470-477), L(478-492), M(493-500), N(504-514), O(517-530), O1(533-542), O2(544-559), P(560-574), 10(591-595), 11(599-603), Q(608-635), 12(645-652), 13(654-661), R(662-685), 14(691-698). 5a, 6a, 6b, 6c, 6d and E' are unique secondary structural elements of the T7 DNA polymerase. Helix J of Klenow fragment is not present in T7 DNA polymerase, in which these residues form a series of irregular turns. The nucleotide bound in the polymerase active site is labeled (cyan). For clarity, only helices α 1- α 4 of thioredoxin (orange) are labelled. **b**, Residues of the polymerase active site. This stereo view shows the polymerase complex in a similar orientation as in **a**. ddGTP and 2 metals (white), primer (magenta), and template (yellow) are shown with the locations of functionally important residues described in the text. **a** and **b** were made with the program SETOR⁴⁸.



the phosphodiester backbone (Fig. 2). Although helix I of other Pol I-type enzymes is interrupted by a bulge or kink at the base of the thumb¹⁴, these residues do not contact the DNA in crystal structures of Klenow fragment¹² or Taq polymerase¹⁶. The 3'–5' exonuclease domain of T7 DNA polymerase does not contact the DNA in this structure.

Roles of conserved residues and metals

DNA polymerases catalyse phosphoryl transfer through the nucleophilic attack of the primer 3'-hydroxyl on the α -phosphate of a dNTP, resulting in nucleotide incorporation and the release of

pyrophosphate⁸. Several conserved sequence motifs have been implicated by biochemical and structural studies of this reaction. Prominent among these is the highly conserved motif that decorates one surface of the O helix of Pol I family polymerases^{4,19}. In the T7 DNA polymerase complex, ddGTP binds in a crevice between the thumb and fingers, forming a Watson–Crick base pair with the template cytosine. The base of the incoming nucleotide stacks between the O helix and the DNA base at the primer 3'-end. The template cytosine is secured in the active site by contacts to its flanking phosphates and a base-stacking interaction with the O helix (Fig. 2). The phosphate 5' of the template cytosine is hydrogen-bonded

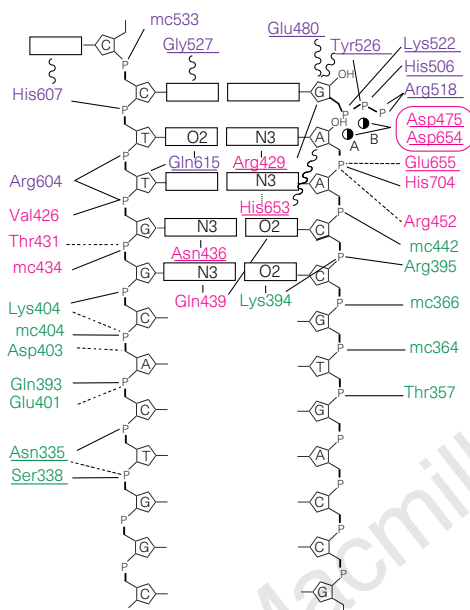


Figure 2 Polymerase contacts to the primer-template. Direct (solid lines) and water-mediated (dashed lines) interactions of T7 DNA polymerase with the primer-template DNA are shown, along with van der Waals packing interactions (wavy lines). Residue labels are colour-coded for the fingers (purple), palm (magenta) and thumb (green) subdomains. Residues with a main-chain contact to the DNA backbone are indicated by the prefix 'mc'. Metals A and B (spheres) contact the phosphates of the nucleoside triphosphate and they are ligated by the strictly conserved carboxylates of Asp 475 and Asp 654. 3'-Hydroxyls are shown for the primer terminus and the bound nucleotide, although they are not present in the crystal structure. The highly conserved residues of the Pol I family are underlined; the corresponding residue numbers for Klenow fragment are: Asn 335 = 579, Ser 338 = 582, Arg 429 = 668, Asn 436 = 675, Asp 475 = 705, Glu 480 = 710, His 506 = 734, Lys 522 = 758, Tyr 526 = Phe 762, Gly 527 = 763, Gln 615 = 849, His 653 = 881, Asp 654 = 882, Glu 655 = 883.

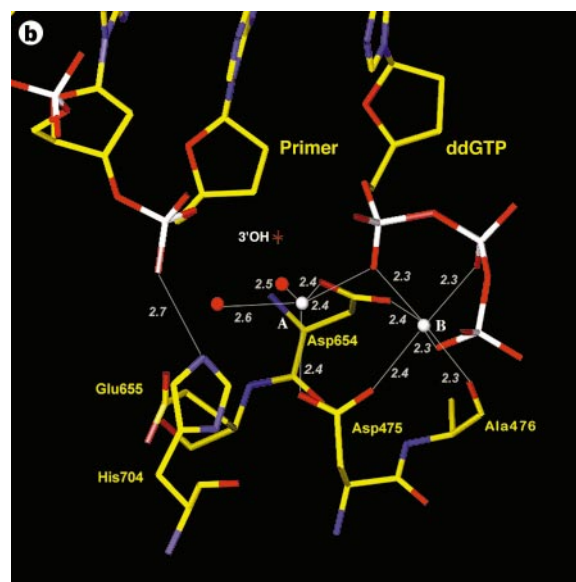
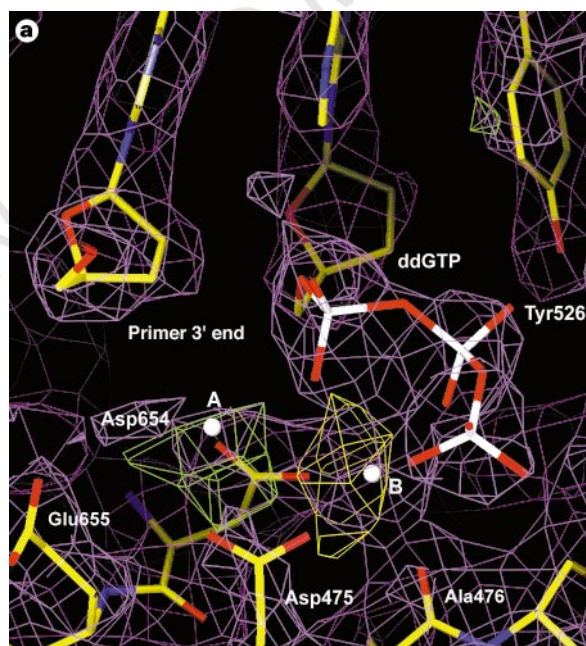


Figure 3 Two-metal ligation of the bound nucleotide. **a**, The 2.5 Å experimental electron density map (purple) was phased by multiple wavelength anomalous diffraction and improved by the program DM⁴⁴. The density is superimposed on the refined model and isomorphous difference Fourier maps showing the Zn²⁺ density (green) and the Mn²⁺ density (yellow). Metals A and B interact with the strictly conserved aspartic acids Asp 475 and Asp 654, and their refined positions are shown as white spheres. Glu 655 does not bind either metal in the crystal

structure. **b**, Two metals in the polymerase active site ligate the unesterified oxygens of all three phosphates of the incoming nucleotide. The interatomic distances (in angstroms) of groups interacting with the metals are shown. The octahedral coordination of metal B is completed by the side chains of Asp 475 and Asp 654, and the main-chain oxygen of Ala 476. Metal A is ligated by Asp 475 and Asp 654, and by two nearby waters. A 3'-hydroxyl (shown by a cross) was modelled on the primer 2',3'-dideoxyribose, and it is located 2.0 Å from metal A.

to the amide nitrogen of Gly 533, and the phosphate on its 3' side is contacted by His 607. The pyrimidine ring of the template cytosine stacks against the C α of a conserved glycine (Gly 527) such that a larger side chain would alter the position of the template base. The nascent base pair between the template and the incoming nucleotide does not form any hydrogen bonds with the protein. Instead, the correct nucleotide is specified by the tight fit of this base pair within the narrow slot between the O helix and the primer terminus. These close-packing interactions restrict the motion of ddGTP, favouring a canonical Watson–Crick base pair with the template. The base step between the incoming nucleotide and the primer 3'-terminus is overwound by more than 10°, an orientation that aligns the primer 3'-hydroxyl for attack of the α -phosphate.

The ribose and triphosphate moieties of the ddGTP are contacted extensively by conserved residues and by two metals that coordinate the unesterified oxygens of all three phosphates. These metals, located 3.6 Å apart, make key interactions that both position the nucleotide and promote phosphoryl transfer. The metals are evident in the experimental electron density map, and their identities have been confirmed by diffraction measurements from crystals soaked in Mn²⁺ or Zn²⁺ (Fig. 3a, Table 1). While it is likely that both metal sites in the native crystals are occupied by Mg²⁺, the experimental electron density at site B is stronger than that of site A. One possible

explanation for this difference is the lack of a hydroxyl group on the primer that might otherwise stabilize the binding of metal A. Site A, which is readily substituted with Zn²⁺, is within 2.0 Å of a 3'-hydroxyl superimposed on the primer terminus. This metal contacts the pro R_p-oxygen of the α -phosphate of ddGTP (Fig. 3b). (The unmodified phosphorus is a prochiral centre. Non-bridging oxygens are designated pro S_p and pro R_p. Substitution with a thiol group makes P_i truly chiral.) Site B, which can bind Mn²⁺, contacts non-bridging oxygens of all three phosphates of ddGTP, including the pro R_p-oxygen of the α - and β -phosphate. Both metals are contacted by the strictly conserved residues Asp 475 and Asp 654 (Fig. 3b). The main-chain oxygen of Ala 476 completes the octahedral coordination of metal B, and two water molecules complete the coordination of metal A (Fig. 3b). The ligation of the pro R_p-oxygen of the α - and β -phosphates by metals is consistent with the selective incorporation of phosphorothioate-containing nucleotide analogues by *E. coli* Pol I. In the presence of Mg²⁺, *E. coli* Pol I is unable to use dATP analogues with sulphur substituted for the pro R_p-oxygen of the α -phosphate or β -phosphate, whereas substitutions involving the pro S_p-oxygen are tolerated²⁰. Mn²⁺ supports the use of β -[R_p]-thio-dATP, consistent with this metal's greater affinity for sulphur. A third highly conserved carboxylate, Glu 655, is located near the metal-ligating residues Asp 475 and Asp 654.

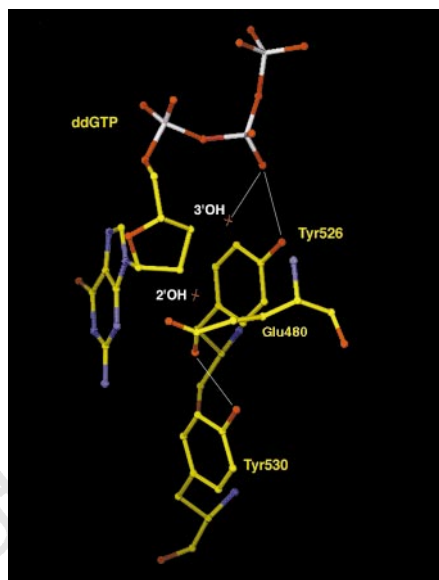


Figure 4 Selectivity for 2'-deoxyribonucleotides. C2' of the incoming ddGTP is wedged between Tyr526 and the aliphatic portion of Asp480. This hydrophobic pocket would not accommodate the 2'-hydroxyl of ribonucleotide substrates (position indicated by a cross). Discrimination at the 3' position of the incoming nucleotide is strongly influenced by the residue at the position of Tyr526. The hydroxyl of Tyr526 is within hydrogen bonding distance of the β -phosphate of ddGTP, as is a 3'-hydroxyl modelled on the ddGTP (cross). One or both of these hydroxyl groups are essential for efficient incorporation of the nucleotide²⁴.

Table 1 Data collection and refinement statistics

Data collection	SeMet-1	SeMet-2	SeMet-3	SeMet-4	Native	MnCl ₂	ZnCl ₂
Wavelength (Å)	0.9788	0.9822	0.9500	1.5418	0.9500	1.5418	1.5418
Resolution limits (Å)	2.2	2.5	2.5	2.5	2.2	3.1	3.5
Total observations	162,012	110,495	123,745	137,675	153,202	55,373	39,476
Unique observations	60,190	41,291	41,938	42,390	58,916	20,903	13,700
<i>R</i> _{sym}	0.054	0.041	0.042	0.058	0.033	0.107	0.123
<i>R</i> _{sym} (last shell)	0.151	0.088	0.093	0.221	0.091	0.323	0.294
<i>I</i> / σ (last shell)	6.5	9.4	9.5	5.2	9.6	3.0	3.2
Completeness	0.965	0.962	0.973	0.989	0.941	0.896	0.868
Completeness (last shell)	0.867	0.858	0.888	0.949	0.892	0.743	0.845
Model refinement							
<i>R</i> _{cryst} / <i>R</i> _{free}	0.240	0.279					
Resolution:	20–2.2 Å						
r.m.s. deviations from stereochemical target values:							
r.m.s. deviation, bond length (Å)	protein: 0.006	DNA: 0.005					
r.m.s. deviation, bond angles (°)	protein: 1.19	DNA: 1.14					
<i>B</i> -factors (Å ²)	protein: 24.2	DNA: 19.4	metals (A,B): 28.5, 20.6		waters: 21.5		

*R*_{sym} = $\sum |h - \langle h \rangle| / \sum |h|$, where $\langle h \rangle$ is the average intensity over symmetry equivalents.

*R*_{cryst} and *R*_{free} = $\sum |F_o| - |F_c| / \sum |F_o|$, where *F*_o and *F*_c are the observed and calculated structure factor amplitudes. *R*_{free} was calculated with 10% of the reflections not used during refinement.

However, in the crystal structure, Glu 655 does not contact either metal, consistent with the modest effect of mutating the analogous residue (Glu 883) in Klenow fragment¹⁰. Glu 655 interacts with the C-terminal residue of the polymerase, His 704, which in turn contacts the phosphate at the primer 3'-end. An altered polymerase in which alanine is substituted for His 704 does not support phage growth, suggesting this interaction with the primer 3'-terminus is essential (S.T. and C.C.R., unpublished results).

The metals in the T7 DNA polymerase active site orient the bound nucleotide and they are likely to assist in the phosphoryl transfer reaction chemistry^{8,21}. Metal A's position in the polymerase active site is compatible with a role in activating the 3'-hydroxyl of the primer, and both metals could counteract the developing negative charge of the transition state. Two metals in the active site of mammalian DNA polymerase- β make similar interactions with a bound nucleotide⁶, including an α , β , γ -tridentate coordination of the phosphates by one of the metals¹⁷. Thus, the essential features proposed for the two-metal mechanism of DNA polymerization⁸ are consistent with crystallographic studies of these two highly divergent enzymes.

A number of conserved residues in the polymerase active site place additional steric and electrostatic constraints on binding the incoming ddGTP. All of the non-bridging phosphate oxygens not contacted by metals interact with one or more conserved residues of the O helix. Two oxygens of the γ -phosphate are contacted by Arg 518. Tyr 526 and His 506 (helix N) contact the β -phosphate, and Lys 522 contacts the α -phosphate (Fig. 2). Mutating any of the analogous residues in Klenow fragment (Arg 754, Phe 762, His 734, Lys 758) markedly diminishes catalytic activity^{10,19,22}. The ribose moiety of the incoming nucleotide is wedged between the aromatic ring of Tyr 526 and the aliphatic carbons of the Glu 480 side chain, which hydrogen-bonds to the hydroxyl of the strictly conserved Tyr 530 (Fig. 4). These residues form a hydrophobic pocket at the C2' position of the ribose that could exclude ribonucleotides from

the polymerase active site, thus providing a structural basis for the strong discrimination against ribonucleotide incorporation by DNA polymerases. Substitution of the Glu 480 analogue in Klenow fragment (Glu 710) with an alanine greatly relaxes discrimination against ribonucleotides²³.

Most polymerases of the Pol I family discriminate strongly against 2',3'-dideoxynucleoside triphosphates (ddNTPs), whereas T7 DNA polymerase readily incorporates these chain-terminating analogues. Tyr 526 in the O helix is a key determinant of efficient use of ddNTPs by Pol I-type enzymes²⁴. A phenylalanine substitution of Tyr 526 in T7 DNA polymerase increases discrimination against ddNTPs several thousand-fold. The reciprocal substitution, a tyrosine for the phenylalanine at the homologous position of Klenow fragment or Taq DNA polymerase, relaxes their discrimination against ddNTPs by several thousand-fold. A 2.5 Å structure of T7 DNA polymerase complexed with dGTP and primer-template (not shown) is essentially identical to the complex presented here. The 3'-hydroxyl of the incoming dGTP and the hydroxyl of Tyr 526 are both within hydrogen-bonding distance of the pro S_p -oxygen of the β -phosphate (Fig. 4). One or both of these interactions with the β -phosphate may be required for efficient incorporation of nucleotides.

Open and closed conformations of the fingers

During synthesis by T7 DNA polymerase, the primer-template is translocated with each cycle of polymerization at the rate of several hundred bases per second²⁵. This movement of the DNA is likely to require a conformational change in the polymerase. The multitude of interactions between T7 DNA polymerase and its bound substrates (Figs 1 and 2) suggest that the protein must relax its grip on the DNA in order for translocation to occur. Crystal structures of Klenow fragment alone²⁶ or with DNA bound in the 3'-5' exonuclease site¹², and Taq DNA polymerase alone¹³ or complexed with DNA¹⁶ show an orientation of the fingers and thumb that is different

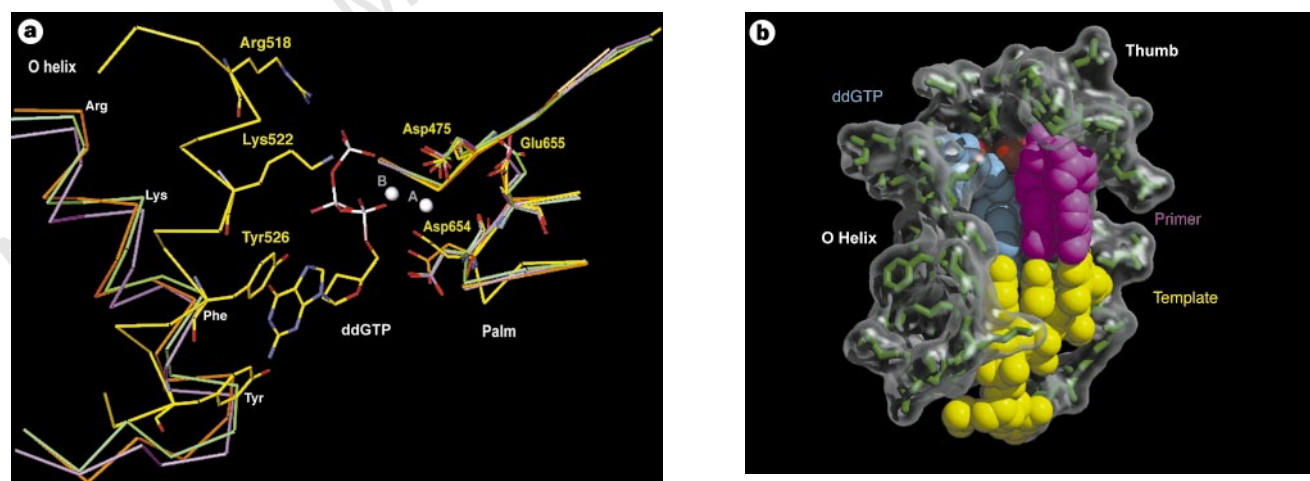


Figure 5 Open and closed conformations of polymerase active sites. **a**, Structures of the T7 DNA polymerase (yellow) and other Pol I family enzymes were superimposed, using the C_{α} coordinates of the highly conserved β -strands 9, 12 and 13 of the palm. This alignment of the Klenow fragment editing complex (orange¹²), the unliganded Taq DNA polymerase (purple¹³; an additional structure of Taq DNA polymerase¹⁵ is superimposable and it is not shown), and the Taq polymerase-DNA complex (green¹⁶) overlays three strictly conserved active site carboxylates of all these polymerases (T7 residues Asp 475, Asp 654 and Glu 655; Klenow fragment residues Asp 705, Asp 882 and Glu 883). For clarity, only the residues of the O helix and a portion of the palm domain are shown, along with the nucleotide and two metals from the T7 DNA polymerase complex. The O helices of Klenow fragment and the two Taq DNA polymerase structures are in an open conformation, whereas the O helix of T7 DNA polymerase is rotated inward

towards the palm. This closed conformation brings the four conserved O helix residues in contact with the incoming nucleotide (T7 residues Arg 518, Lys 522, Tyr 526 and Tyr 530; Klenow fragment residues Arg 754, Lys 758, Phe 762 and Tyr 766). **b**, A space-filling model of the primer-template and the incoming nucleoside triphosphate is shown with the accessible surface of nearby residues displayed as a transparent surface. Two bound metals are partially visible (red spheres). The base of the ddGTP (cyan) stacks against the adenine at the 3'-end of the primer (magenta), and it forms a Watson-Crick base pair with the template (yellow). This nascent base pair between the template and the incoming nucleotide lies in a narrow slot that disfavors the binding of mismatched or staggered base pairs. A cytidine located to the 5' side of the template base is flipped out of the active site and it stacks on the surface of the fingers (bottom of figure).

from that of the T7 DNA polymerase complex. In all of these structures the fingers are in the same open conformation, with the O helix rotated away from the metal-ligating carboxylates of the palm (Fig. 5a). In contrast, the fingers subdomain of T7 DNA polymerase complexed to primer–template and ddGTP is rotated inwards by about 41° towards the primer–template, such that the O helix abuts the nucleotide-binding site (Fig. 5b). This closed orientation of the fingers allows the functionally important O helix residues Arg 518, Lys 522 and Tyr 526 to contact the sugar and all three phosphates of the incoming nucleotide. The opened and closed orientations of the fingers subdomains seen in crystal structures of various Pol I family polymerases might correspond to different states in the reaction pathway. The rate-limiting step in the catalytic cycle of Klenow fragment²⁷ and T7 DNA polymerase²⁵ occurs after the dNTP binds and before it is incorporated. Originally posited as a change in polymerase conformation, this slow step may correspond to the closure of the fingers around the bound nucleotide in Pol I-type enzymes. Similarly, DNA polymerase-β adopts a closed conformation when the correct dNTP forms a base pair with the template in the polymerase active site¹⁷.

Misincorporation and proofreading

DNA synthesis slows tremendously following misincorporation of a nucleotide²⁸, allowing excision of the incorrect nucleotide by the proofreading exonuclease²⁹. A reduced rate of DNA synthesis might result from the weaker binding of a primer–template containing a mismatch, or its misalignment with respect to the polymerase active site. T7 DNA polymerase contacts a limited number of DNA bases including the base pair located at the 3′-end of the primer (Figs 2 and 6). Arg 429 donates a hydrogen bond to N3 of adenine at the primer 3′-terminus, and Gln 615 donates a hydrogen bond to O2 of thymine in the complementary strand. These interactions between two strictly conserved residues of the Pol I family and the ‘universal’ hydrogen bond acceptors of the minor groove³⁰ are at key positions for detecting a misincorporated base. Alanine substitutions of the analogous residues in Klenow fragment (Arg 668, Gln 849) markedly decrease DNA-binding affinity and the k_{cat} of DNA synthesis^{9,10}. Mismatched base pairs are accommodated in duplex DNA by the rotation of one or both bases into the major groove³¹, which would disrupt minor groove interactions with the polymerase and possibly trigger the switch from DNA synthesis to proofreading.

The structure of the 3′–5′ exonuclease domain of T7 DNA polymerase (Fig. 1) is similar to that of Klenow fragment¹². Alignment of the palms of these polymerases superimposes their exonuclease active sites. The active site of the T7 3′–5′ exonuclease is located at the base of a deep cleft where four conserved acidic residues ligate two metals located 4.5 Å apart (not shown). One metal-binding site is readily substituted by Mn²⁺ and the other by

Zn²⁺, in agreement with the metal-binding preferences of the 3′–5′ exonuclease active site of Klenow fragment²⁰. Both metals in the T7 exonuclease domain are chelated by Asp 5 and, in addition, the metal in the Zn²⁺ site is bound by Glu 7 and Asp 174. Asp 65 makes water-mediated interactions with the metal in the Mn²⁺ site. Metals at similar positions in the exonuclease site of Klenow fragment engage the bridging phosphate of a dinucleotide substrate, where they could assist in phosphodiester bond cleavage²⁰. In the crystal structure of the Klenow fragment editing complex, the 3′-end of the primer strand bends sharply downward into the exonuclease active site¹², a path that is blocked in T7 DNA polymerase by the segment between helices D and F (Fig. 1a). This obstacle might be an artefact caused by the deletion of six residues (118–123) within this region of the exonuclease-deficient polymerase. Alternatively, DNA substrates may enter the exonuclease site through a more exposed path located near the base of the thumb.

Processivity domain of the polymerase

High processivity is an important attribute of all replicative DNA polymerases. Some achieve processive synthesis by associating with a multimeric protein ring termed a sliding clamp, which encircles the primer–template duplex³². T7 DNA polymerase uses the comparatively simple strategy of binding to *E. coli* thioredoxin in a stable one-to-one complex that has a processivity rivaling that of more complicated enzymes¹. Thioredoxin binds to an extended loop located between helices H and H1 of the thumb (Fig. 1a). This 71-residue loop, which is not present in other Pol I-type polymerases, wraps around the base of thioredoxin and buries the active site cysteines (Cys 32, Cys 35) as well as the Arg 73–Gly 74–Ile 75–Pro 76 motif. The proximity of Gly 74 of thioredoxin and Glu 319 of the polymerase was previously suggested on the basis of genetic and biochemical studies³³. The thioredoxin-binding loop is notably rich in glycine, proline and lysine residues, and the average temperature factor of atoms within the loop is nearly twice that of the rest of the protein (Table 1), indicating increased mobility or flexure. Twenty-five residues at the distal end of the loop are disordered in the crystal structure, including three functionally important lysines that affect phage growth and polymerase activity *in vitro*³⁴. The structure of reduced thioredoxin in the polymerase complex is nearly identical to that of oxidized thioredoxin alone (PDB accession number 2TRX; r.m.s deviation = 0.86 Å for Cα atoms 4–107), yet oxidized thioredoxin does not bind to the polymerase³⁵. Moreover, the replacement of either cysteine in thioredoxin with serine or alanine reduces its affinity for T7 polymerase³⁶. Cys 32 of thioredoxin forms a hydrogen bond with Thr 327, effectively decreasing its polarity within the hydrophobic subunit interface and contributing to the requirement for reduced thioredoxin. The extended segment around Thr 327 lies between Cys 32 and Pro 76 of thioredoxin, in

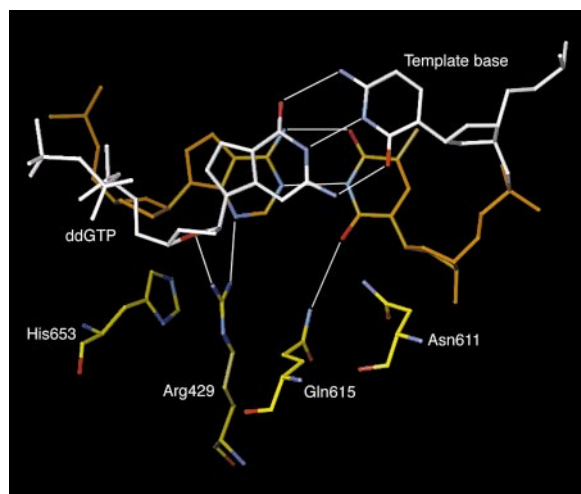


Figure 6 Misincorporation and proofreading. The nascent base pair in the polymerase active site (white) and the adjacent base pair of the primer–template (orange) are shown along with four strictly conserved residues located on the floor of the active site in this region. Arg 429 and Gln 615 contact the minor groove N2 and O3 positions of the base pair at the primer 3′-end. A mismatch at this position, resulting from misincorporation by the polymerase, might be sensed as a loss of one or both of these minor groove interactions. Figures 3–6 were made with the program O⁴⁵, except for Fig. 5b which was generated with GRASP⁴⁹ and Raster3D⁵⁰.

the same groove occupied by a NF- κ B peptide with human thioredoxin³⁷.

Although thioredoxin is rotated away from the primer–template in the crystal structure, this appendage of the thumb prevents the dissociation of a primer–template during DNA synthesis¹. Thioredoxin and the associated binding loop appear to be flexibly tethered to the thumb (Fig. 1a), and they could swing across the DNA-binding groove to encircle the primer–template exiting the polymerase. Consistent with this hypothesis, an additional seven base pairs of the primer–template are protected from exonuclease digestion upon the addition of thioredoxin (S.T. and C.C.R., unpublished data). The processivity domain could secure the primer–template in the polymerase, either by forming a lid over the active site that functions like a sliding clamp, or by making electrostatic interactions with the DNA. Insertion of this domain into the thumb of the Klenow fragment polymerase confers a thioredoxin-dependent increase in processive DNA synthesis³⁸, suggesting that the processivity mechanism of T7 DNA polymerase is largely contained within this compact domain.

This study of the T7 DNA polymerase complexed with its processivity factor, a primer–template, and nucleoside triphosphate provides the first structure of a replicative DNA polymerase locked in a polymerization mode. The complex shows an unprecedented closed conformation of the polymerase active site that allows catalytically important, conserved residues to engage the incoming nucleotide. Two aspartate residues that are strictly conserved in the Pol I family, and that have homologues in other DNA polymerases and in RNA polymerases, ligate two metals in the polymerase active site in positions that are consistent with a two-metal mechanism of phosphoryl transfer. This crystal structure will guide additional studies of the accessory proteins that work in concert with the polymerase during leading and lagging strand synthesis, leading ultimately to a high-resolution picture of the T7 replisome. □

Methods

Crystallization and data collection. T7 DNA polymerase (Δ 118–123) was overproduced with *E. coli* thioredoxin from a T7 gene 10 promoter in BL21-pLysS *E. coli* and purified to homogeneity in the presence of thioredoxin using three chromatographic steps: phosphocellulose (Whatman), anion exchange Poros HQ (PerSeptive Biosystems), and ceramic hydroxylapatite (BioRad). Selenomethionyl T7 DNA polymerase (Δ 118–123) was produced in *E. coli*³⁹, taking care to maintain strictly reducing conditions during purification. Synthetic oligonucleotides were purified by anion exchange HPLC (Poros HQ medium). The sequence of the template strand in the crystals is 5'-CCCCTTGGCACTGGCCGTCGTTTTTCG-3', and that of the primer is 5'-CGAAAACGACGGCCAGTGCCA-3'. These strands were annealed and mixed in a one-to-one ratio with the T7 DNA polymerase–thioredoxin heterodimer, in the presence of a chain terminator (ddATP) and excess of the next nucleotide (ddGTP). Crystals of the T7 DNA polymerase complex were obtained by using an incomplete factorial design⁴⁰. Orthorhombic crystals ($P2_12_12_1$; $a = 106.6$ Å, $b = 216.3$ Å and $c = 52.1$ Å) grew overnight after microseeding against a reservoir solution made of 12% (w/v) polyethylene glycol 8000, 100 mM ACES pH 7.5, 120 mM ammonium sulphate, 30 mM MgCl₂, and 5 mM dithiothreitol. Our initial crystals of the complex with dGTP deteriorated over the course of several days. Crystal decay coincided with the hydrolysis of dGTP to dGMP, apparently due to an idling reaction at the 3'-end of the template involving the polymerase and residual proofreading exonuclease. Substitution of dGTP with ddGTP prevented nucleotide hydrolysis and resulted in the growth of larger, well-ordered crystals. Before data collection, crystals were stabilized in a solution containing all the reservoir components plus polyethylene glycol 400, then mounted in a nylon loop and flash-frozen in liquid propane. These crystals contain one polymerase–DNA–nucleotide complex ($M_r = 108$ K) in the asymmetric unit and they diffract X-rays to better than 2.2 Å at 100 K using synchrotron radiation. Native data and multiple wavelength anomalous diffraction (MAD) data from a selenomethionine-containing crystal (SeMet 1–3) were collected at beamline X12C of the National Synchrotron Light Source (Upton, NY) using a MAR detector. The mono-

chromator positions were based on X-ray absorption fluorescence of the selenomethionine-containing crystal, and one additional data set was collected from the same crystal using a laboratory X-ray source. Following a 6-h soak in MnCl₂ (5 mM) or ZnCl₂ (1 mM) the crystallographic unit cell was isomorphous to that of native frozen crystals, but diffraction quality decreased (Table 1).

Phasing and refinement. All X-ray intensity data were processed with DENZO/SCALEPACK⁴¹ and scaled using local scaling (MAXSCALE)⁴². The positions of all 15 seleniums were located by direct methods (SHELXS-86)⁴³, using the dispersive differences between the inflection point and the high-energy remote data (SeMet-2 minus SeMet-3). Heavy-atom parameters were refined and phases computed with MLPHARE⁴⁴. The resulting experimental MAD map was of excellent quality, with clear solvent boundary. This map was further improved by histogram matching and solvent flattening using the program DM⁴⁴. Model building and adjustments were done with the program O⁴⁵ first building into the experimental MAD map, then into SIGMAA weighted maps⁴⁴. The positions of the 15 methionines guided the assignment of residues in the electron density. The model of the polymerase complex was refined with XPLOR version 3.851⁴⁶. The current model includes 766 residues (residues 4–107 for thioredoxin, 1–293, 318–575, and 587–704 for T7 DNA polymerase Δ 118–123), 24 DNA bases, one molecule of ddGTP, and 526 water molecules with refined temperature factors of less than 40 Å². There are two *cis*-prolines in the complex: Pro 76 (thioredoxin) and Pro 435 (T7 DNA polymerase). Only one of 766 residues has (ϕ , ψ) angles in a disallowed region. This residue, His 653 ($\phi = 69.4$, $\psi = 79.3$), precedes the conserved acidic residues Asp 654 and Glu 655. A similar strained conformation has been reported for the equivalent residue in the structure of *Bacillus stearothermophilus* DNA polymerase fragment¹⁴. Coordinates of the T7 DNA polymerase complex have been deposited in the Brookhaven Protein Databank (accession number 1T7P).

Received 11 July; accepted 29 September 1997.

1. Tabor, S., Huber, H. E. & Richardson, C. C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262**, 16212–16223 (1987).
2. Modrich, P. & Richardson, C. C. Bacteriophage T7 deoxyribonucleic acid replication *in vitro*. A protein of *Escherichia coli* required for bacteriophage T7 DNA polymerase activity. *J. Biol. Chem.* **250**, 5508–5514 (1975).
3. Debyser, Z., Tabor, S. & Richardson, C. C. Coordination of leading and lagging strand DNA synthesis at the replication fork of bacteriophage T7. *Cell* **77**, 157–166 (1994).
4. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461–467 (1990).
5. Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787–802 (1993).
6. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, 1891–1903 (1994).
7. Tabor, S. & Richardson, C. C. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis. *J. Biol. Chem.* **264**, 6447–6458 (1989).
8. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).
9. Polesky, A. H., Steitz, T. A., Grindley, N. D. & Joyce, C. M. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **265**, 14579–14591 (1990).
10. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. & Joyce, C. M. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **267**, 8417–8428 (1992).
11. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762–766 (1985).
12. Beese, L. S., Derbyshire, V. & Steitz, T. A. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**, 352–355 (1993).
13. Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E. & Waksman, G. Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-Å resolution: structural basis for thermostability. *Proc. Natl Acad. Sci. USA* **92**, 9264–9268 (1995).
14. Kiefer, J. R. *et al.* Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95–108 (1997).
15. Kim, Y. *et al.* Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376**, 612–616 (1995).
16. Eom, S. H., Wang, J. & Steitz, T. A. Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **382**, 278–281 (1996).
17. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism. *Biochemistry* **36**, 11205–11215 (1997).
18. Minnick, D. T., Astatke, M., Joyce, C. M. & Kunkel, T. A. A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *J. Biol. Chem.* **271**, 24954–24961 (1996).
19. Astatke, M., Grindley, N. D. & Joyce, C. M. Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **270**, 1945–1954 (1995).
20. Burgers, P. M. & Eckstein, F. A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. *J. Biol. Chem.* **254**, 6889–6893 (1979).
21. Beese, L. S. & Steitz, T. A. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33 (1991).

22. Kaushik, N., Pandey, V. N. & Modak, M. J. Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis: dynamics of the dNTP binding pocket. *Biochemistry* **35**, 7256–7266 (1996).
23. Joyce, C. M. Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc. Natl Acad. Sci. USA* **94**, 1619–1622 (1997).
24. Tabor, S. & Richardson, C. C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc. Natl Acad. Sci. USA* **92**, 6339–6343 (1995).
25. Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511–525 (1991).
26. Beese, L. S., Friedman, J. M. & Steitz, T. A. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32**, 14095–14101 (1993).
27. Bryant, F. R., Johnson, K. A. & Benkovic, S. J. Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry* **22**, 3537–3546 (1983).
28. Wong, I., Patel, S. S. & Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526–537 (1991).
29. Donlin, M. J., Patel, S. S. & Johnson, K. A. Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* **30**, 538–546 (1991).
30. Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804–808 (1976).
31. Hunter, W. N., Brown, T., Anand, N. N. & Kennard, O. Structure of adenine-cytosine base pair in DNA and its implications for mismatch repair. *Nature* **320**, 552–555 (1986).
32. Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425–437 (1992).
33. Himawan, J. S. & Richardson, C. C. Amino acid residues critical for the interaction between bacteriophage T7 DNA polymerase and *Escherichia coli* thioredoxin. *J. Biol. Chem.* **271**, 19999–20008 (1996).
34. Yang, X. M. & Richardson, C. C. Amino acid changes in a unique sequence of bacteriophage T7 DNA polymerase alter the processivity of nucleotide polymerization. *J. Biol. Chem.* **272**, 6599–6606 (1997).
35. Adler, S. & Modrich, P. T7-induced DNA polymerase. Requirement for thioredoxin sulphydryl groups. *J. Biol. Chem.* **258**, 6956–6962 (1983).
36. Huber, H. E., Russel, M., Model, P. & Richardson, C. C. Interaction of mutant thioredoxins of *Escherichia coli* with the gene 5 protein of phage T7. The redox capacity of thioredoxin is not required for stimulation of DNA polymerase activity. *J. Biol. Chem.* **261**, 15006–15012 (1986).
37. Qin, J., Clore, G. M., Kennedy, W. M., Huth, J. R. & Gronenborn, A. M. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B. *Structure* **3**, 289–297 (1995).
38. Bedford, E., Tabor, S. & Richardson, C. C. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proc. Natl Acad. Sci. USA* **94**, 479–484 (1997).
39. Doublé, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
40. Carter, C. W. Jr in *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (eds Ducruix, A. & Giegé, R.) 47–71 (IRL Press, New York, 1992).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
42. Rould, M. A. Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Methods Enzymol.* **276**, 461–472 (1997).
43. Sheldrick, G. M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467–473 (1990).
44. Bailey, S. The CCP4 Suite—programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, I. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
46. Brünger, A. T. XPLOR Version 3.1: A system for X-ray crystallography and NMR. (Yale University Press, New Haven, CT, 1992).
47. Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
48. Evans, S. V. SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).
49. Nicholls, A., Chapp, K. A. & Honig, B. Protein folding and association: insights from interfacial and thermodynamics properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
50. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **276**, 505–524 (1997).

Acknowledgements. We thank R. M. Sweet and F. W. Studier for access and assistance with the use of beamline X12C (NSLS, Brookhaven National Laboratory). We thank H. J. Kwon, A. Lau and M. Rould for experimental help at the synchrotron, members of the Ellenberger, Richardson and Hogle groups for many discussions and M. Rould for programs and discussions. This work was supported by the National Institutes of Health, the Lucille Markey Charitable Trust, the Department of Energy, the Harvard Center for Structural Biology and the Armenise-Harvard Foundation for Advanced Scientific Research. C.C.R. and S.T. are consultants to Amersham Life Science Inc., which has licenses from Harvard University to commercialize DNA polymerases for use in DNA sequencing.

Correspondence and requests for materials should be addressed to T.E. (e-mail: tome@jeckle.med.harvard.edu) and S.T. (e-mail: stabor@heckle.med.harvard.edu).

KNOW YOUR COPY RIGHTS RESPECT OURS

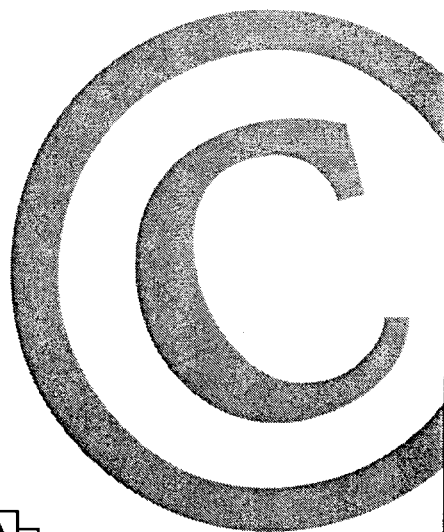
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

The interstellar chemistry of PAH cations

Theodore P. Snow*, Valery Le Page†, Yeghis Keheyan‡ & Veronica M. Bierbaum†

* Center for Astrophysics and Space Astronomy, University of Colorado, Campus Box 389, Boulder, Colorado 80309, USA

† Department of Chemistry and Biochemistry, University of Colorado, Campus Box 215, Boulder, Colorado 80309, USA

‡ Istituto di Chimica Nucleare del CNR, Monterotondo Stazione, 00016 Roma, Italy

Diffuse interstellar bands (DIBs) are mysterious absorption lines in the optical spectra of stars, and have been known for 75 years¹. Although it is widely believed^{2–4} that they arise from gas-phase organic molecules (rather than from dust grains) in the interstellar medium, no consensus has been reached regarding their precise cause. The realization that many emission features in astronomical infrared spectra probably arise from polycyclic aromatic hydrocarbons^{5–8} (PAHs), which may themselves be very abundant in the interstellar medium, has led to the suggestion that ionized PAHs might be the source of the DIBs^{9–12}. Laboratory investigations have revealed that small, positively charged PAHs in matrices have absorption features that bear some resemblance to DIBs^{13–15}, but no clear identification of any DIB with any specific PAH cation has yet been made. Here we report a laboratory study of the chemical reactivity of PAH cations ($C_6H_6^+$, $C_{10}H_8^+$ and $C_{16}H_{10}^+$) in the gas phase. We find that these PAH cations are very reactive, and are therefore unlikely to survive in high abundances in the interstellar medium. Rather, such molecules will react rapidly with hydrogen, and we therefore suggest that the resulting protonated PAH cations (and species derived from them) should become the focus of future searches for a correspondence between molecular absorption features and the DIBs.

The experiments were carried out in a tandem-flowing afterglow–selected ion flow tube (FA–SIFT)¹⁶ which is ideally suited to the study of reactions between ions and atoms. The $C_6H_6^+$, $C_{10}H_8^+$ and $C_{16}H_{10}^+$ ions were formed by Penning ionization of the neutral parent species with metastable argon atoms which are generated in a cold cathode discharge¹⁷; this relatively gentle ionization process cleanly generates the parent cation. The dehydrogenated ions were formed by chemical ionization with helium ions that were produced by electron impact ionization. In both cases the reactant ions were extracted from the source region, mass selected, and injected into the reaction flow tube where they were thermalized by collisions with helium. We studied the dehydrogenated forms $C_{10}H_7^+$ and $C_{16}H_9^+$ because in the diffuse interstellar medium such species are likely to be present as a result of photodissociation of the parent PAH cations.

The selection of neutral reactants was guided by the expected abundance in the diffuse interstellar medium¹⁸. The DIBs seem to form most efficiently in H I clouds, which are dominated by atomic hydrogen but which also contain substantial quantities of molecular hydrogen, atomic nitrogen and atomic oxygen (other cosmically abundant elements such as carbon are predominantly ionized in H I clouds and will not react with PAH cations).

In our experiments H_2 was added through a manifold of inlets, and kinetics were measured as a function of reaction distance. Atomic hydrogen was generated by passing ultra-high-purity H_2 through a microwave discharge tube; calibration reactions were used to measure the resulting atomic hydrogen density^{19–21}, and indicated that dissociation ratios were about 30%. Atomic nitrogen

was formed by passing N_2 through the microwave discharge, and atomic oxygen was formed by titration of atomic nitrogen with nitric oxide²²: $N + NO \rightarrow N_2 + O$. Dissociation ratios of N_2 were about 2%. Small corrections to the rate constants were made to account for wall recombination and mixing of the atomic reagents.

We also studied the reactions of the dehydrogenated PAH cations with molecular hydrogen at low pressure with a dual-cell, 3-tesla Fourier transform–ion cyclotron resonance instrument (Extrel FTMS 2001)²³. For these experiments $C_{10}H_7^+$ was formed both by electron impact on bromonaphthalene and by a displacement reaction of SiF_3^+ with fluoronaphthalene.

Table 1 lists our results for the reactions of the benzene^{19,20,24}, naphthalene and pyrene parent cations, and their dehydrogenated and protonated forms, with the most common collision partners in the diffuse interstellar medium (reactions with other species such as simple molecules are reported elsewhere²⁵). The high rates for reactions of the parent PAH cations with H I, combined with the overwhelming predominance of H I over any other atomic or molecular collision partners in the diffuse interstellar medium, suggests that the chemistry of these cations will be dominated by reactions with H I. These hydrogenation reactions of $C_6H_6^+$, $C_{10}H_8^+$ and $C_{16}H_{10}^+$ produce protonated forms ($C_6H_7^+$, $C_{10}H_9^+$ and $C_{16}H_{11}^+$). Similarly, the dehydrogenated ions ($C_6H_5^+$, $C_{10}H_7^+$ and $C_{16}H_9^+$) react rapidly with molecular hydrogen or in two successive reactions with atomic hydrogen to form the protonated species. These protonated ions react with atomic hydrogen but with much smaller rate coefficients ($\sim 4 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$).

Our laboratory results for the chemical reactivity suggest that the protonated forms are likely to dominate in the diffuse interstellar medium where the DIBs arise. However, there are important competing physical processes (such as electron recombination, photoionization and photodissociation) which must be considered; detailed modelling is required for a full characterization of PAH ion densities in the interstellar medium.

We can offer a few comments, however. For example, under typical diffuse-cloud conditions, number density of hydrogen atoms $n_H = 10 \text{ cm}^{-3}$ and temperature $T = 100 \text{ K}$ (these are conditions under which pyrene is expected to be about 50% ionized¹²) we can determine the hydrogenation timescale for PAH cations, $\tau_H = 1/kn_H$. We assume that k , the rate coefficient for radiative association with atomic hydrogen, is roughly 42% of the measured rate coefficient at 0.5 torr helium pressure (in analogy with the $C_{10}H_7^+ + H_2$ reaction, Table 1) and that the rate coefficient has a $T^{-3/2}$ dependence. Thus for ionized pyrene, $k = 3.1 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ and $\tau_H = 10.2 \text{ yr}$; therefore, the pyrene cation has a relatively short lifetime before reacting with a hydrogen atom. Similar considerations show that ionized naphthalene will react with atomic hydrogen ($k = 4.2 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$) on a timescale of 7.6 yr. These timescales will decrease at higher densities, increasing the hydrogenation rate.

It is perhaps more interesting to compare timescales for hydrogenation ($\tau_H = 1/kn_H$) and electron recombination ($\tau_e = 1/\alpha n_e$), the major competing physical process¹². Here α is the electron recombination rate coefficient and n_e is the electron density. The ratio of the two timescales is $\tau_H/\tau_e = \alpha n_e/kn_H$. It is typically assumed that all of the electrons in the diffuse interstellar medium come from the ionization of carbon, so that the ratio n_e/n_H equals the cosmic ratio of carbon to hydrogen, which we take²⁶ to be 1.4×10^{-4} ; this ratio is independent of the overall cloud density. Thus the ratio of timescales becomes $1.4 \times 10^{-4}(\alpha/k)$. The coefficient α has been measured only for the benzene and naphthalene cations²⁷; here, for illustrative purposes, we consider the latter ($C_{10}H_8^+$). Using the value $k = 4.2 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ as discussed above, and taking^{12,27} $\alpha(T) = 3.0 \times 10^{-7}(T/300)^{-1/2}$, we find that at $T = 100 \text{ K}$ the ratio of timescales is 0.17. Thus in a typical diffuse cloud hydrogenation dominates (by a factor of 6) over recombination, and we can expect PAH cations to migrate to their protonated form before

Table 1 Reactions between PAH cations and major interstellar species

Reactants	H ₂	H	O	N
Benzene				
C ₆ H ₆ ⁺	5.0 × 10 ⁻¹¹	—	—	—
C ₆ H ₆ ⁺	no reaction	2.2 × 10 ⁻¹⁰	9.5 × 10 ⁻¹¹	1.2 × 10 ⁻¹⁰
C ₆ H ₇ ⁺	no reaction	~4 × 10 ⁻¹²	—	—
Naphthalene				
C ₁₀ H ₇ ⁺	5.2 × 10 ⁻¹¹	—	—	—
C ₁₀ H ₈ ⁺	no reaction	1.9 × 10 ⁻¹⁰	1.0 × 10 ⁻¹⁰	2.3 × 10 ⁻¹¹
C ₁₀ H ₉ ⁺	no reaction	~4 × 10 ⁻¹²	—	—
Pyrene				
C ₁₆ H ₉ ⁺	no reaction	~1.6 × 10 ⁻¹⁰	~2 × 10 ⁻¹⁰	~3 × 10 ⁻¹¹
C ₁₆ H ₁₀ ⁺	no reaction	1.4 × 10 ⁻¹⁰	9.5 × 10 ⁻¹¹	1.5 × 10 ⁻¹²
C ₁₆ H ₁₁ ⁺	no reaction	~3 × 10 ⁻¹²	—	—

All rate constants have been measured in a SIFT apparatus at 0.5 torr and 300 K, with the exception of C₆H₆⁺ + H₂ (refs 19, 24), C₆H₆⁺ + H (ref. 19, 20) and C₆H₆⁺ + H (refs 19 and 20, weighted average). The rate coefficients (units of cm³ s⁻¹) are accurate to +100–50% for reactions between protonated species (C₆H₆⁺, C₁₀H₈⁺ and C₁₆H₁₁⁺) and H atoms, ±50% for reactions involving N atoms, and ±30% otherwise. The dashes indicate very low reactivity and only upper limits were determined; the blanks indicate that the reactions were not studied. Reactions between PAH cations and H/H₂ lead primarily to association whereas for O and N, formation of CO and HCN, respectively, is also seen. The dehydrogenated species C₆H₅⁺, C₁₀H₇⁺ and C₁₆H₉⁺ are mixtures of two isomers; the table gives the overall reactivity with atoms and the reactivity of only the cyclic form with H₂. The other reactant ions are cyclic. We have also measured the purely radiative association rate constant (2.2 × 10⁻¹¹ cm³ s⁻¹) for the reaction C₆H₇⁺ + H₂ in a FTICR at low pressure at 300 K. The low reactivity of C₁₆H₁₁⁺ toward H₂ and the high reactivity towards H are understood as a consequence of a triplet ground state for this ion; this ion was also found to be unreactive with H₂ in our FTICR experiments.

recombining. (If the radiative association rate constant (k) for the reaction of ionized naphthalene with atomic hydrogen is assumed to equal that for the reaction of C₁₀H₇⁺ with molecular hydrogen, and k is assumed to have a $T^{-3/2}$ dependence, $\tau_H = 27.7$ yr and $\tau_H/\tau_e = 0.64$. Thus, even in this extreme case, hydrogenation dominates over recombination by a factor of nearly 2.)

There are several important unknowns in the recombination of PAH cations. Comprehensive model calculations require accurate values of α for larger parent cations, as well as for the protonated forms. In addition, the relative contribution of the dissociative and non-dissociative channels is critical; these processes can directly form radical neutral species which may have visible absorption spectra, or closed-shell neutral PAHs which can be photoionized to radical cations. The recycling between neutral and ionic forms, as well as radical and non-radical species, must be fully modelled.

The question of PAH ionization balance is crucial to the hypothesis that the DIBs are formed by PAH cations. Although it is true that the DIBs do not form efficiently in dense clouds where the PAHs are expected to be either neutral or negatively charged²⁸, it is also true that they are weak in regions having very high ultraviolet fluxes²⁹. However, small protonated PAH cations are not expected to have visible absorption spectra; thus if small PAHs contribute to the DIBs, they are likely to be more fully hydrogenated radical cations (such as C₁₀H₁₀⁺ and C₁₆H₁₂⁺) or radicals formed by electron recombination processes.

There are good reasons to consider large protonated PAH cations as the carriers of the DIBs. Unlike small protonated forms, larger protonated cations will absorb in the visible. The trends which we have observed in chemical reactivity³⁰ suggest that such cations will also be important in the interstellar medium and therefore may directly contribute to the DIBs. The larger species should also be less prone to photodissociation. Small neutral PAHs have substantial photodissociation rates³¹, and even though no rate data exist for the ionized forms, it is likely that the same would be true of small PAH cations. But large species, neutral or ionized, will have sufficient internal energy modes to allow photons to be absorbed non-destructively³¹.

In view of this, it is tempting to hypothesize that the weakness of the DIBs in areas of high ultraviolet flux²⁹ might be due to photodestruction of the carriers. This could explain why the DIBs

are strong in diffuse environments where there is sufficient ultraviolet flux to ionize PAHs, but weak in regions of more intense ultraviolet radiation, where photodissociation would be enhanced. It has been pointed out that PAH anions might also be considered as DIB carriers¹², but the weakness of the DIBs in dense or dark clouds²⁸, where anions are most likely to dominate^{12,32}, argues against this suggestion.

Our results should guide future studies of spectroscopy aimed at testing the PAH⁺ hypothesis. We suggest that measurements of spectra be concentrated on hydrogenated forms rather than only the parent PAH cations. It will be important to attempt measurements of large PAH cations as well. In addition, more detailed experimental studies of the rates and products of electron recombination and photodissociation reactions are required for modelling of the chemical and physical processes that determine the distribution and densities of PAH ions and neutrals in the interstellar medium. □

Received 11 August; accepted 11 November 1997.

- Heger, M. L. The spectra of certain class B stars in the regions 5630 Å–6680 Å and 3280 Å–3380 Å. *Lick Obs. Bull.* **10**, 146–147 (1922).
- Tielens, A. G. G. M. & Snow, T. P. (eds) *The Diffuse Interstellar Bands* (Kluwer, Dordrecht, 1995).
- Herbig, G. H. The diffuse interstellar bands. *Annu. Rev. Astron. Astrophys.* **33**, 19–73 (1995).
- Snow, T. P. in *From Stardust to Planetesimals* (eds Pendleton, Y. J. & Tielens, A. G. G. M.) 147–157 (Astron. Soc. Pacific, San Francisco, 1997).
- Léger, A. & Puget, J. Identification of the “unidentified” IR emission features of interstellar dust? *Astron. Astrophys.* **137**, L5–L8 (1984).
- Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. Polycyclic aromatic hydrocarbons and the unidentified infrared emission bands: auto exhaust along the Milky Way! *Astrophys. J.* **290**, L25–L28 (1985).
- Puget, J. & Léger, A. A new component of the interstellar matter: small grains and large aromatic molecules. *Annu. Rev. Astron. Astrophys.* **27**, 161–198 (1989).
- Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. Interstellar polycyclic aromatic hydrocarbons: the infrared emission bands, the excitation/emission mechanism, and the astrophysical implications. *Astrophys. J. Suppl.* **71**, 733–775 (1989).
- Crawford, M. K., Tielens, A. G. G. M. & Allamandola, L. J. Ionized polycyclic aromatic hydrocarbons and the diffuse interstellar bands. *Astrophys. J.* **293**, L45–L48 (1985).
- Léger, A. & d’Hendecourt, L. Are polycyclic aromatic hydrocarbons the carriers of the diffuse interstellar bands in the visible? *Astron. Astrophys.* **146**, 81–85 (1985).
- van der Zwet, G. P. & Allamandola, L. J. Polycyclic aromatic hydrocarbons and the diffuse interstellar bands. *Astron. Astrophys.* **146**, 76–80 (1985).
- Salama, F., Bakes, E. L. O., Allamandola, L. J. & Tielens, A. G. G. M. Assessment of the polycyclic aromatic hydrocarbon—Diffuse interstellar band proposal. *Astrophys. J.* **458**, 621–636 (1996).
- Salama, F. & Allamandola, L. J. The ultraviolet and visible spectrum of the polycyclic aromatic hydrocarbon C₁₀H₈: Possible contributions to the diffuse interstellar bands and to the ultraviolet-visible extinction. *Astrophys. J.* **395**, 301–306 (1992).
- Salama, F. & Allamandola, L. J. Is a pyrene-like molecular ion the cause of the 4430 Å diffuse interstellar absorption band? *Nature* **358**, 42–43 (1992).
- Léger, A., d’Hendecourt, L. & Defourneau, D. Proposed identification for the (common) carriers of the 4430 Å and 7565 Å DIBs. *Astron. Astrophys.* **293**, L53 (1995).
- Van Doren, J. M., Barlow, S. E., DePuy, C. H. & Bierbaum, V. M. The tandem flowing-afterglow-SIFT-drift. *Int. J. Mass Spectrom. Ion Processes* **81**, 85–100 (1987).
- Howorka, F. Reactions of singly and doubly charged argon ions with N₂ and O₂ in a steady state hollow cathode discharge. *J. Chem. Phys.* **68**, 804–811 (1978).
- Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble Space Telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–329 (1996).
- Petrie, S., Javahery, G. & Bohme, D. K. Gas-phase reactions of benzenoid hydrocarbons and ions with hydrogen atoms and molecules: Uncommon constraints to reactivity. *J. Am. Chem. Soc.* **114**, 9205–9206 (1992).
- Scott, G. B. I. et al. C_mH_n⁺ reactions with H and H₂: An experimental study. *J. Phys. Chem.* **101**, 4973–4978 (1997).
- Tosi, P. et al. The reaction of CO₂⁺ with atomic hydrogen. *J. Chem. Phys.* **80**, 1905–1906 (1984).
- Setser, D. W. (ed.) *Reactive Intermediates in the Gas Phase* (Academic, New York, 1979).
- de Petris, G., Occhiucci, G. & Pepi, F. A mass spectrometric study of gaseous H₄PO₃⁺ and H₂PO₃⁺ ions. *Int. J. Mass Spectrom. Ion Processes* **136**, 155–166 (1994).
- Giles, K., Adams, N. G. & Smith, D. A study of reactions of C_mH_n⁺ ions ($n = 4, 5, 6$; $m = 0–6$) with H₂ and CO at 300 K and 80 K. *Int. J. Mass Spectrom. Ion Processes* **89**, 303–317 (1989).
- Le Page, V., Keheyan, Y., Bierbaum, V. M. & Snow, T. P. Chemical constraints on organic cations in the interstellar medium. *J. Am. Chem. Soc.* **119**, 8373–8374 (1997).
- Snow, T. P. & Witt, A. N. The interstellar carbon budget and the role of carbon in dust and large molecules. *Science* **270**, 1455–1460 (1995).
- Abouelaziz, H., Gomet, J. C., Pasquerault, D. & Rowe, B. R. Measurements of C₃H₃⁺, C₃H₄⁺, C₆H₆⁺, C₇H₇⁺ and C₁₀H₈⁺ dissociative recombination rate coefficients. *J. Chem. Phys.* **99**, 237–243 (1993).
- Adamson, A. J., Whittett, D. C. B. & Duley, W. W. Diffuse interstellar bands in the Taurus dark clouds. *Mon. Not. R. Astron. Soc.* **252**, 234–245 (1991).
- Snow, T. P., Bakes, E. L. O., Buss, R. H. & Seab, C. G. The weakness of diffuse bands in nebular environments: Possible impact on the PAH⁺ hypothesis. *Astron. Astrophys.* **296**, L37–L40 (1995).
- Le Page, V., Bierbaum, V. M., Keheyan, Y. & Snow, T. P. A study of reactions of cations derived from naphthalene with molecules and atoms of interstellar interest. *J. Am. Chem. Soc.* (submitted).
- Allain, T., Leach, S. & Sedlmayr, E. Photodestruction of PAHs in the interstellar medium. I. Photodissociation rates for the loss of an acetylenic group. *Astron. Astrophys.* **305**, 602–615 (1996).
- Omont, A. *Astron. Astrophys.* **164**, 159–178 (1986).

Acknowledgements. This research has been supported by a NASA grant to the University of Colorado.

Correspondence should be addressed to T. P. Snow (e-mail: tsnow@casa.colorado.edu).

Tellurium in pre-solar diamonds as an indicator for rapid separation of supernova ejecta

S. Richter*, U. Ott & F. Begemann

Max-Planck-Institut für Chemie (Otto-Hahn-Institut), Becherweg 27, D-55128 Mainz, Germany

Carbon-rich 'carbonaceous' meteorites contain several types of dust grains with an isotopic signature that identifies them as being of pre-solar origin¹⁻³. Of these grains, diamonds are of particular interest: such grains are by far the most abundant, and they host an isotopically anomalous 'Xe-H' component (characterized by a relative overabundance of the heaviest stable isotopes of xenon) which constitutes a notable fraction of the total amount of xenon in unprocessed 'primitive' meteorites. The isotope abundance ratios of this Xe-H cannot be accounted for by the canonical processes responsible for nucleosynthesis of the elements heavier than iron. An *ad hoc* neutron-capture process has been postulated⁴⁻⁶ to explain the observed isotope abundance ratios, but it has also been pointed out that standard 'r-process' nucleosynthesis (in supernovae) could work if the stable isotopes were somehow separated from their radioactive precursors in the first few hours after the explosion⁷. One way to distinguish between these mechanisms is to determine anomalies correlated for the heavy stable isotopes of tellurium in pre-solar diamond grains. Here we report such measurements, which support the suggestion that the isotopes were separated: the competing neutron-capture process cannot produce the observed abundances.

The anomalous Xe-H found in pre-solar diamond is characterized by large overabundances, relative to Solar-System xenon, of the two heaviest stable isotopes, ¹³⁴Xe and ¹³⁶Xe. As these isotopes generally are thought to have been produced exclusively by the rapid neutron-capture process (r-process)⁸ of nucleosynthesis (where 'rapid' refers to the timescale of successive neutron captures by the same target nuclei), a straightforward explanation for their overabundance is that, in Xe-H, the mixing ratio is higher than 'normal' between r-process Xe and those isotopes that originate from different processes of nucleosynthesis. If this were so, and if there were only a single kind of r-process Xe, then the relative overabundances of ¹³⁴Xe and ¹³⁶Xe would have to be the same. This is not the case; the overabundance at ¹³⁴Xe is only about 60% of that at ¹³⁶Xe (ref. 7). As a way out of this difficulty, a special process has been invented that produces r-Xe with a lower ¹³⁴Xe/¹³⁶Xe ratio than does the average r-process. This is the 'mini r-process' or neutron burst, which can be tailored to produce xenon with the required ¹³⁴Xe/¹³⁶Xe ratio; it is thought to occur either in explosive carbon burning, or in the outer He-rich mantle of an exploding supernova when the shock wave traverses it⁶.

Despite more than 30 years having elapsed since the discovery⁹ of anomalous Xe-H, the match is still poor between the predictions of this type of model^{5,6} and the observed pattern^{7,10}. This, and the intellectual dissatisfaction of having to resort to a special process to account for an uncomfortably large fraction of total Xe (in the Allende meteorite, for example, Xe-H amounts to about 8% of total ¹³⁶Xe), has led to the development of an alternative, the 'rapid-separation' model⁷. This model uses the fact that in the r-process of nucleosynthesis stable isotopes are not produced directly but

through the radioactive decay of their respective isobaric precursors¹¹. Because the half-lives in the different decay chains are different, the abundance ratios of the stable end-products change with time. In the particular case of the ¹³⁴Xe/¹³⁶Xe ratio it asymptotically increases with time towards its final value. As has been shown⁷, if one values the standard r-process and separates out Xe from its precursors after a few hours at this intermediate stage (that is before it approaches its final value), and one adds a small amount of fully developed r-process Xe (to account for finite contributions at ^{129,131,132}Xe), the calculation provides a perfect match to the observed Xe-H pattern).

Among the possibilities for achieving a separation are the precondensation of refractory elements which will remove longer-lived

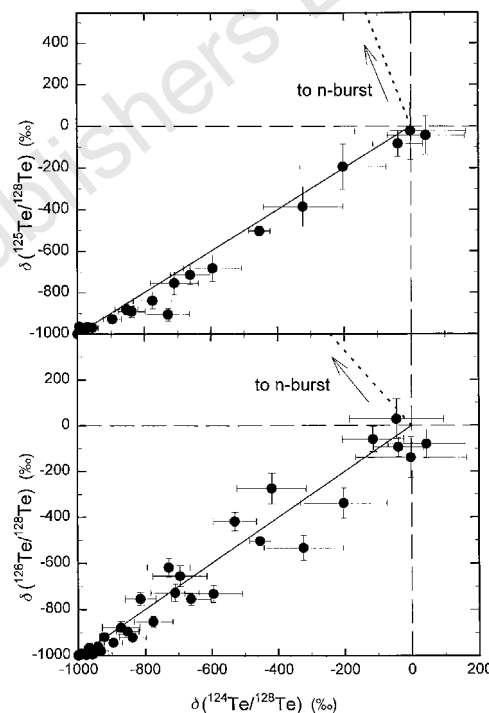


Figure 1 Tellurium isotopic composition measured in pre-solar diamonds extracted from the Allende meteorite. Data are shown as δ -values (deviation from norm¹⁵ in per mil) in three-isotope plots of ¹²⁵Te/¹²⁸Te (top) and ¹²⁶Te/¹²⁸Te (bottom) against ¹²⁴Te/¹²⁸Te. The residue was prepared following established procedures¹⁹. Te analyses were performed by thermal ionization mass spectrometry of negative Te ions (N-TIMS), using the double-filament technique and the multichannel-ion-counting MAT 262 (MIC-MAT 262) mass spectrometer at Mainz²⁰. Several splits, ~10 μ g each, of the diamond residue were loaded for analysis as a suspension in H₂O/HNO₃ (pH $\approx$ 1) together with 1 μ l of saturated H₃BO₃ on the evaporation filament. Ba(OH)₂ as emitter²¹ and H₃BO₃ were loaded on the ionization filament. Ratios changed during the course of the measurements from very anomalous (lower left) to more normal values; the individual data points shown are ratios calculated from the number of ions collected in intervals of 80 s duration. Because of the low count rates (often <1 ion per second for the less-abundant isotopes), the precision in the ratios is essentially determined by ion statistics; uncertainties (2 σ) are assigned according to Poisson statistics. Data points fall along mixing lines between normal Te (at 0, 0) and a component containing ¹²⁸Te, but virtually no ^{124,125,126}Te (at -1,000, -1,000). As will be discussed elsewhere, we have satisfied ourselves that the latter component is not an artefact from isobaric and/or molecular interferences. The composition of this anomalous component is consistent with the prediction for Te-H from the 'early loss' variant of the rapid-separation model⁷. In contrast, mixing of normal Te with Te as predicted from the neutron burst model⁶ would produce compositions lying along the dashed lines marked 'to n-burst'.

* Present address: Institute for Reference Materials and Measurements, Joint Research Centre, Commission of the European Communities, B-2440 Geel, Belgium.

precursor nuclides from the reservoir before they have completely decayed into the respective stable xenon isotopes⁷. Alternatively, the separation might occur by the early trapping of r-process-produced matter within tiny, nanometre-sized (possibly pre-existing) grains, followed by subsequent loss from these grains by recoil resulting from the radioactive decay of such precursor nuclides that were sufficiently long-lived to decay within the grains⁷. It is unclear whether any of the mechanisms considered so far will work in a supernova environment on the required timescale. Given that dust formation is observed to occur only after about a year¹², rapid separation through precondensation must be considered highly unlikely, although it might be helped by the inhomogeneities^{13,14} observed within supernova ejecta.

To check whether it agrees with the predictions of one of the models we have measured the isotopic composition of tellurium in pre-solar diamonds from the Allende meteorite that are known to contain Xe-H. Tellurium was chosen because its isotopic structure is similar to that of its neighbour, xenon, in that both elements possess isotopes believed to originate from a single process of nucleosynthesis: ¹²⁰Te from the p-process, ^{122–124}Te solely from the slow

neutron capture s-process, and ^{128,130}Te solely from the r-process (plus ^{125,126}Te of mixed s- and r-origin). In the present context the existence of the two r-process-only isotopes is of particular importance because it is the relative abundance in Xe-H of their xenon counterparts, ¹³⁴Xe and ¹³⁶Xe, that poses a major problem.

Relevant results are shown, as δ -values (deviations from the norm¹⁵ in per mil) in the three-isotope plots of Figs 1 and 2. The data in Fig. 1, for isotopes 124, 125, 126 and 128, are suggestive of mixing between Te of normal isotopic composition (all δ -values $\equiv$ 0) and Te with an extremely anomalous composition, where the anomalous Te is essentially free from ^{124,125,126}Te ($\delta = -1000$ ‰). (The same holds true for ^{120,122,123}Te (not shown).) As to the heavy isotopes, the ordinate intercept in Fig. 2 suggests the anomalous ¹³⁰Te/¹²⁸Te abundance ratio to be about 8% smaller than normal ($\delta = -80$ ‰). In summary, the inferred composition of the anomalous Te in pre-solar diamonds from the Allende carbonaceous chondrite is ¹²⁰Te : ¹²²Te : ¹²³Te : ¹²⁴Te : ¹²⁵Te : ¹²⁶Te : ¹²⁸Te : ¹³⁰Te = $< 4 \times 10^{-5} : < 0.0012 : < 0.0008 : \equiv 0 : 0.0006 \pm 0.0018 : 0.0001 \pm 0.0037 : 1.008 \pm 0.018 : \equiv 1$ (Fig. 3). It consists essentially only of the two heaviest stable isotopes, ¹²⁸Te and ¹³⁰Te; consequently we designate it Te-H.

The absence from Te-H of ¹²⁰Te (p-only isotope) and ^{122,123,124}Te (s-only isotopes) is gratifying but of little diagnostic value, as this is predicted by both the neutron burst and all variants of the rapid-separation model. The absence of ¹²⁵Te and ¹²⁶Te, however, is highly significant because it rules out the neutron burst model, which predicts a sizable production of both isotopes⁶ (Fig. 3). Actually, only one of the existing models to explain the isotopic composition of Xe-H, as we know it from pre-solar diamonds, can also account for the basic features of Te-H from the same pre-solar diamonds. This is the variant where rapid trapping is followed by the loss due to recoil of decay products ('early loss' variant). It predicts the presence of ¹²⁸Te and ¹³⁰Te as well as the absence of ¹²⁵Te and ¹²⁶Te because these are shielded by long-lived isobaric precursors, ¹²⁵Te by ¹²⁵Sb (half-life, $T_{1/2} = 2.77$ yr), and ¹²⁶Te by ¹²⁶Sn with a half-life of about 10^5 years. (The variant of the model requiring precondensation of all elements but the most volatile ones, like xenon, fails because it predicts the absence of the radioactive precursors of r-process Te as well as that of any early-formed stable Te; in other words, it does not allow for any kind of Te-H.)

The 'early loss' variant of the rapid separation model also has a problem, however, which is the exact value of the ¹²⁸Te/¹³⁰Te ratio (Fig. 3). The different timescales in building up the final abundances

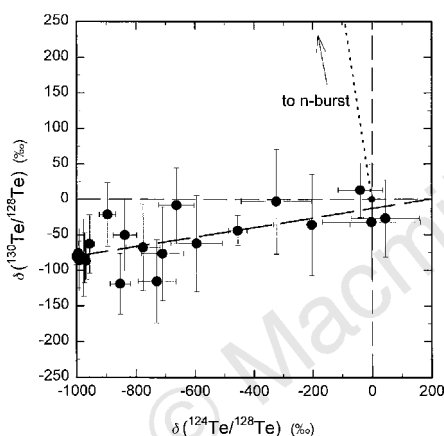


Figure 2 Three-isotope plot of ¹³⁰Te/¹²⁸Te against ¹²⁴Te/¹²⁸Te. Data points fall along mixing lines between normal Te (at 0, 0) and a Te component with ¹³⁰Te/¹²⁸Te about 8% lower than normal. The 8% ¹³⁰Te deficit may be indicative of complexity in the r-process (see text for further discussion).

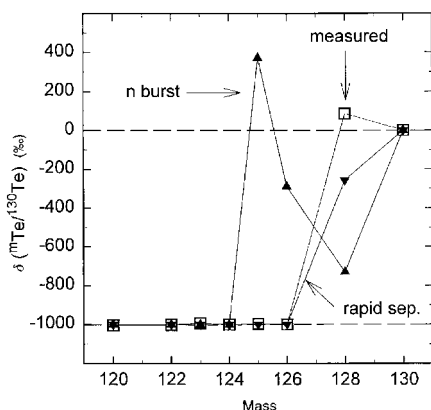


Figure 3 Comparison of isotopic composition of Te-H, as derived here, with predictions of models. Models considered are the neutron burst⁶ model and the 'early loss' variant of the rapid-separation model⁷. Plotted are δ -values, which are deviations from the normal terrestrial composition.

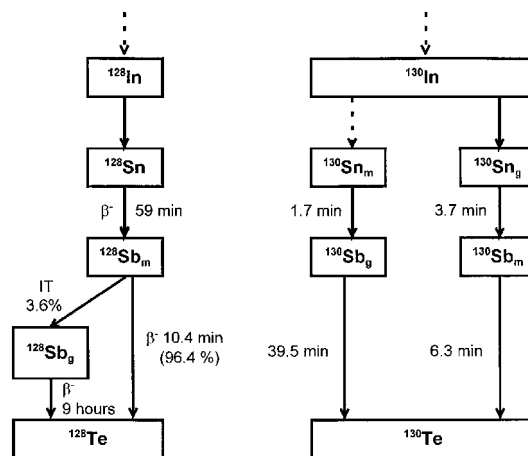


Figure 4 Schemes for the decay of radioactive r-process precursors into stable ¹²⁸Te and ¹³⁰Te (ref. 22). Although the exact timescale for the production of ¹³⁰Te depends on how the ¹³⁰In states are filled, in any case its final abundance is established faster than that of ¹²⁸Te, where the timescale is defined by the decay of ¹²⁸Sn (59 min). g, ground state; m, isomeric state; IT, internal transition.

of the stable isotopes after termination of the r-process are such that the final abundance of ^{130}Te is established faster than that of ^{128}Te (Fig. 4). Hence, at any given time, $^{128}\text{Te}/^{130}\text{Te}$ can only be lower than, or equal to, the 'production ratio' (the abundance ratio after complete decay). As the model assumes that this production ratio is that in the 'average' r-process, which, in turn, is equal to that used as a reference in Fig. 3, deviations from this value can only be negative. What we find, however, is a ratio that is 8% higher than the production ratio.

The only way to get around this problem within the framework of the rapid-separation model is to relax the assumption of an 'average r-process' source. Although this may seem *ad hoc*, it may not be unreasonable, as there are indications of a component structure in the r-process^{16,17} similar to that found for the s-process¹⁸. The r-process runs on a timescale of seconds, and, depending on actual time, the resulting abundances are differently distributed among the abundance peaks associated with the various magic neutron numbers¹⁶. Different running times may be associated with different types of supernovae sources and, based on the relative abundances in the early Solar System of the various extinct radionuclides of r-process origin, the case has been made for diverse supernovae to have been the sources for the matter in the different abundance peaks¹⁷. The hypothesis that a specific type of supernova source is responsible for Xe-H and Te-H is amenable to testing, for example by the analysis in the diamonds of elements belonging to the $A \approx 195$ abundance peak associated with magic neutron number 126.

In evaluating the evidence for or against the competing models, it should be kept in mind that the neutron burst model of ref. 6 was constructed *ad hoc* with the aim of reproducing the observed Xe-H pattern, which includes not only ^{134}Xe and ^{136}Xe , but also finite abundances of ^{129}Xe , ^{131}Xe and ^{132}Xe . It has, however, been shown that the observed abundances in Xe-H of these isotopes can be accounted for by a small admixture of fully developed r-process Xe to a pure ($^{134}\text{Xe} + ^{136}\text{Xe}$) component⁷. Hence, it may be worthwhile to attempt a fresh neutron burst approach, trying to find a solution that produces only ^{134}Xe and ^{136}Xe . If such a solution exists, it might also solve the case of the tellurium isotopes. Independent of this, efforts should be undertaken to determine whether a mechanism exists that would separate radioactive precursors from stable end-products on the timescale of hours required by the rapid separation model. □

Received 21 July; accepted 4 November 1997.

- Ott, U. Interstellar grains in meteorites. *Nature* **364**, 25–33 (1993).
- Anders, E. & Zinner, E. Interstellar grains in primitive meteorites: diamond, silicon carbide, and graphite. *Meteoritics* **28**, 490–514 (1993).
- Zinner, E. Presolar material in meteorites: an overview. *Proc. Astrophysical Implications of the Laboratory Study of Presolar Materials* (AIP Conf. Ser., in the press).
- Heymann, D. & Dziczkaniec, M. Xenon from intermediate zones of supernovae. *Proc. Lunar Planet. Sci. Conf.* **X**, 1943–1959 (1979).
- Clayton, D. D. Origin of heavy xenon in meteoritic diamonds. *Astrophys. J.* **340**, 613–619 (1989).
- Howard, W. M., Meyer, B. S. & Clayton, D. D. Heavy-element abundances from a neutron burst that produces Xe-H. *Meteoritics* **27**, 404–412 (1992).
- Ott, U. Interstellar diamond xenon and time scales of supernova ejecta. *Astrophys. J.* **463**, 344–348 (1996).
- Beer, H., Corvi, F. & Mutti, P. Neutron capture of the bottleneck isotopes ^{138}Ba and ^{208}Pb , s-process studies, and the r-process abundance distribution. *Astrophys. J.* **474**, 843–861 (1997).
- Reynolds, J. H. & Turner, G. Rare gases in the chondrite Renazzo. *J. Geophys. Res.* **69**, 3263–3281 (1964).
- Huss, G. R. & Lewis, R. S. Noble gases in presolar diamonds I: Three distinct components and their implications for diamond origins. *Meteoritics* **29**, 791–810 (1994).
- Clayton, D. D. *Principles of Stellar Evolution and Nucleosynthesis* (Univ. Chicago Press, 1983).
- Kosaka, T., Hasegawa, H. & Nomoto, K. Formation of dust grains in the ejecta of SN 1987A. II. *Astron. Astrophys.* **249**, 474–482 (1991).
- Haas, M. R. et al. Velocity-resolved far-infrared spectra of [Fe II]: evidence for mixing and clumping in SN 1987A. *Astrophys. J.* **360**, 257–266 (1990).
- Chevalier, R. A. Supernova 1987A at five years of age. *Nature* **355**, 691–696 (1992).
- Wachsmann, M. & Heumann, K. G. Negative thermal ionization mass spectrometry of main group elements Part 2. 6th group. *Int. J. Mass. Spectrom. Ion Proc.* **114**, 209–220 (1992).
- Kratz, K.-L., Bitouzet, J.-P., Thielemann, F.-K., Möller, P. & Pfeiffer, B. Isotopic r-process abundances and nuclear structure far from stability: implications for the r-process mechanism. *Astrophys. J.* **403**, 216–238 (1993).
- Wasserburg, G. J., Busso, M. & Gallino, R. Abundances of actinides and short-lived nonactinides in the interstellar medium: diverse supernova sources for the r-process. *Astrophys. J.* **466**, L109–L112 (1996).
- Käppeler, F., Beer, H. & Wisshak, K. S-process nucleosynthesis—nuclear physics and the classical model *Rep. Progr. Phys.* **52**, 945–1013 (1989).

- Amari, S., Lewis, R. S. & Anders, E. Interstellar grains in meteorites: I. Isolation of SiC, graphite, and diamond; size distributions of SiC and graphite. *Geochim. Cosmochim. Acta* **58**, 459–470 (1994).
- Richter, S., Ott, U. & Begemann, F. Multiple ion counting in isotope abundance mass spectrometry. *Int. J. Mass. Spectrom. Ion Proc.* **136**, 91–100 (1994).
- Heumann, K. G., Schindlmeier, W., Zeiningner, H. & Schmidt, M. Application of an economical and small thermal ionization mass spectrometer for accurate anion trace analyses. *Fresenius Z. Anal. Chem.* **320**, 457–462 (1985).
- Firestone, R. B. *Table of Isotopes* (Wiley, New York, 1996).

Correspondence should be addressed to U.O. (e-mail: ott@mpch-mainz.mpg.de).

Quantum interference in electron collision

R. C. Liu*, B. Odom*, Y. Yamamoto*† & S. Tarucha†

*ERATO Quantum Fluctuation Project, Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA

†NTT Basic Research Laboratories, 3-1 Morinosato-Wakamiya Atsugi, Kanagawa, 243-01, Japan

The indistinguishability of identical quantum particles can lead to quantum interferences that profoundly affect their scattering^{1,2}. If two particles collide and scatter, the process that results in the detection of the first particle in one direction and the second particle in another direction interferes quantum mechanically with the physically indistinguishable process where the roles of the particles are reversed. For bosons such as photons, a constructive interference between probability amplitudes can enhance the probability, relative to classical expectations, that both are detected in the same direction—this is known as 'bunching'. But for fermions such as electrons, a destructive interference should suppress this probability ('anti-bunching'); this interference is the origin of the Pauli exclusion principle, which states that two electrons can never occupy the same state. Although two-particle interferences have been shown for colliding photons^{3,4}, no similar demonstration for electrons exists^{2,5,6}. Here we report the realization of this destructive quantum interference in the collision of electrons at a beam splitter. In our experiments, the quantum interference responsible for the Pauli exclusion principle is manifest as the suppression in electron current noise after collision.

The scattering of individual particles at a beam splitter is a stochastic process that introduces fluctuations (partition noise) in the output flux, depending on the transmission probability, T (Fig. 1a). If N identical particles are independently scattered in series, a binomial distribution results for the number of particles transmitted to one of the outputs, N_{out} , and the normalized variance (Fano factor) is equal to $\langle \Delta N_{\text{out}}^2 \rangle / \langle N_{\text{out}} \rangle = (1 - T)$. In the limit of an extremely small T , the partition noise approaches the Poisson limit (full shot noise). This single-particle partition process is independent of the quantum statistics of incident particles.

This no longer holds in the case of collisions (Fig. 1b and c). Assuming a beam splitter with $T = 1/2$, the collision of classical particles (independent partitioning of the two inputs) results in probabilities of 1/4 for two particles to be scattered to the left or two to the right, whereas the probability that one particle is scattered into each output is 1/2. If the incident particle fluxes do not fluctuate, each output flux exhibits one-half shot noise.

In the quantum mechanical case, because particles are described by probability amplitudes, when the wavefunctions of two identical particles initially on the left and right overlap, it becomes impossible to distinguish one particle from the other. Applying the postulates of quantum mechanics can then in principle yield results that depend on whether the state is described mathematically by $|1 : \psi_L; 2 : \psi_R\rangle$ (particle (1) on the left and particle (2) on the right) or by $|1 : \psi_R; 2 : \psi_L\rangle$ (vice versa). Only the symmetric and

antisymmetric combinations of these states, $|\psi_{\pm}\rangle = (1/\sqrt{2})[|1 : \psi_L; 2 : \psi_R\rangle \pm |1 : \psi_R; 2 : \psi_L\rangle]$, produce real measurement results. Particles having overall symmetric wavefunctions are called bosons, and those having antisymmetric wavefunctions are called fermions. The photon is a classical example of a boson, whereas the electron is a classic example of a fermion.

The probability amplitudes for the three possible collision outcomes are derived from the evolution of the two-particle wavefunction under the influence of the beam splitter scattering matrix (unitary evolution). The symmetrization or antisymmetrization of the wavefunction yields two contributions to the probability amplitude, a direct and an exchange term. If we consider the output state where both particles are scattered to the left, then both of these terms have the same magnitude and sign (Fig. 1d). For bosons, these terms add to give a probability of 1/2. For fermions, these terms subtract to give completely destructive interference; two fermions never scatter into the same state, a manifestation of the Pauli exclusion principle.

If we consider the output state where one particle is in each output, then, because the unitary evolution introduces a minus sign for reflection from the right input to the right output in order to satisfy power conservation, the direct and exchange terms have the same magnitude but opposite signs (Fig. 1e). Completely destructive interference is now obtained for bosons, whereas the probability for fermions equals one. In this ideal beam splitter, whereas two bosons always scatter into the same port (Fig. 1b) so that the fluctuations in the output flux exhibit full shot noise (although not poissonian noise), one fermion always scatters into each port (Fig. 1c) so that the fluctuations in the output flux are completely suppressed. The fact that quantum interferences profoundly affect the collision noise suggests that a measurement of the output noise should confirm the quantum statistics of the particles.

To observe this, we fabricated a mesoscopic electron beam splitter by electron beam lithography on a GaAs high-mobility, two-

dimensional electron gas system (Fig. 2a). The two input and two output ports are defined by point contact constrictions formed by an etched trench and Schottky gates that deplete underlying electrons when negative gate biases are applied. A 40-nm gate finger down the centre of the scattering region allows the splitting ratio of the beam splitter to be tuned over a narrow range around 50 : 50 partitioning of transmitted electrons. The transport through the device is primarily ballistic, as the length scales for inelastic phonon scattering and elastic ionized impurity scattering at an operating temperature of 1.6 K are much longer than the device size. To improve the signal-to-noise ratio in the noise measurement, an a.c. modulation scheme^{5,7} is used with a cryogenic cascode configuration preamplifier.

The use of ballistic point contacts for the inputs is necessary to achieve streams of single-mode electron wavepackets incident on

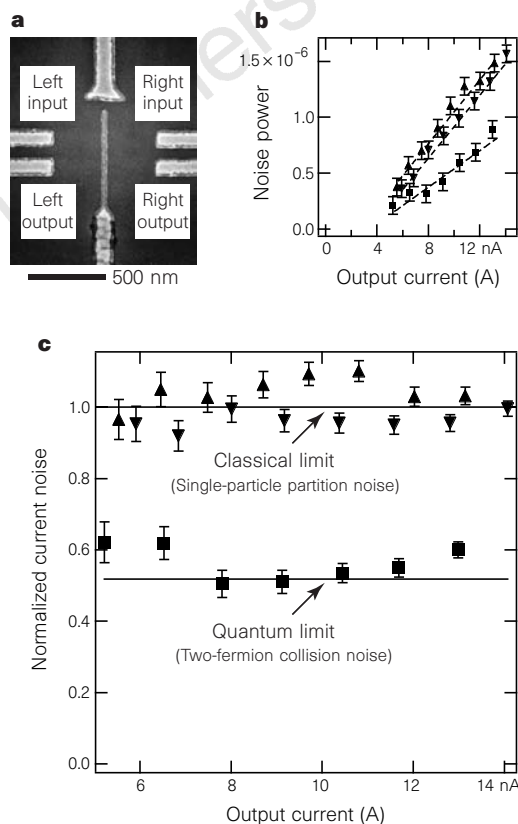


Figure 2 Suppression of collision noise in an electron beam splitter. **a**, Scanning electron micrograph of an electron beam splitter device fabricated on a GaAs two-dimensional electron gas system. Schottky gates and an etched trench (dark area near the centre gate) define the inputs (gate–gate point contacts) and the outputs (gate–trench point contacts). **b**, The magnitude of the current noise power for the right output (arbitrary units), as a function of its current. A modulation scheme and resonant circuit are used to improve the discrimination between the stationary background amplifier noise and the current-dependent partition noise, and thus improve the signal-to-noise ratio at the 15.6 MHz measurement frequency. When either the left input (downward triangles) or the right input (upward triangles) is biased individually, the single-particle partition noises should be and are roughly the same, as the transmission probabilities into the right output are similar. Their slopes are used as references to compare the noise during collision (squares), which is clearly suppressed. **c**, After normalizing the noise by the current (after subtracting zero bias offsets) and scaling by the classical collision noise determined from the weighted average of the slopes of the single-particle partition noises, the measured collision noise can be compared to the classical limit. Its suppression indicates a fermion interference. The noise is not completely suppressed, in part because of non-idealities in the beam splitter's scattering matrix.

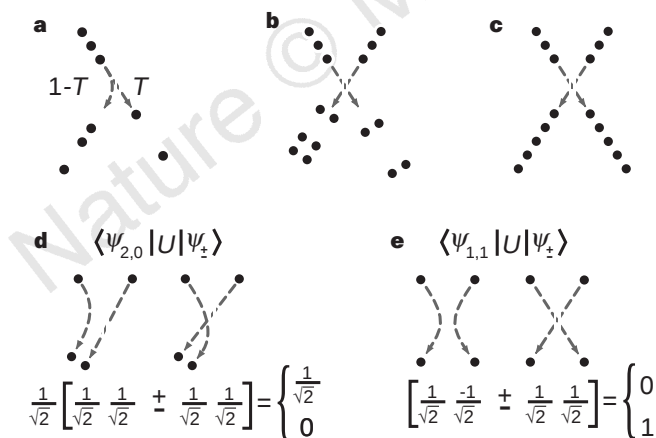


Figure 1 Scattering and collisions at a beam splitter. **a**, An ideal beam splitter with transmission T stochastically partitions a constant incident flux into two output ports. **b**, Collision (for $T = 1/2$) of quantum particles having a symmetric two-particle wavefunction (bosons) results in two particles always in one port or the other. **c**, Collision of quantum particles having an antisymmetric wavefunction (fermions) always results in one particle in each port. **d**, The symmetrized (upper sign) or antisymmetrized (lower sign) input state $|\psi_{\pm}\rangle$ is propagated by the unitary transformation, U , derived from the beam splitter's scattering matrix and projected onto the output state $\langle\psi_{2,0}|$ to give the probability amplitude for the case where two particles are in the left port. This results in a constructive interference between the direct and exchange terms for bosons, and a destructive interference for fermions. **e**, For the output state where one particle is in each port, $\langle\psi_{1,1}|$, because of the additional minus sign due to reflection from the right input to the right output, a destructive interference results for bosons, and a constructive interference arises for fermions.

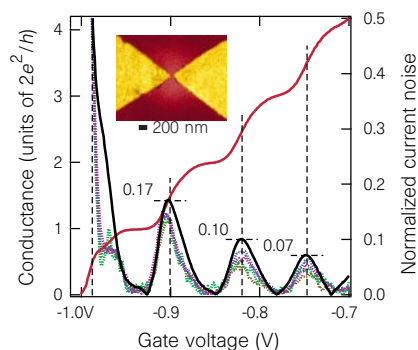


Figure 3 Conductance and partition noise in a quantum point contact. Inset, scanning electron micrograph of an isolated quantum point contact device. Main figure, conductance (solid, red) and current noise normalized by the full shot noise value (dotted) (scaled to the expected Fano factor of $1/2$ at $T = 1/2$ in the first plateau), as a function of the voltage applied to the confining Schottky gates. The current through a point contact is carried by discrete transverse modes. As the point contact width is increased with decreasing negative gate bias, the transmission probability of a mode increases from zero to unity as its transverse sub-band energy drops below the Fermi energy. Once fully transmitting ($T = 1$), a transverse mode carries a fixed current per unit voltage corresponding to G_0 , and a plateau is reached in the conductance. The partition noise is absent at these plateaux. The current noise is measured for four different source-drain bias voltages of 0.5 (green), 0.75 (brown), 1.25 (blue) and 1.75 mV (violet). The noise power peaks at the middle of the steps where $T = 1/2$ for the newly opening mode. The theoretical behaviour of the partition noise, based on the measured conductance assuming all fully opened modes have $T = 1$, is also plotted (solid, black) and shows that the trend in experimental noise peaks agrees with theoretical prediction.

the beam splitter. Both the conductance^{8,9} and noise properties^{7,10,11} of such a device have been actively studied. For our purposes, measurements on an isolated point contact (Fig. 3 inset) serve to illustrate the expected noise contribution of the Schottky gate point contacts used as inputs in the collision experiment. The conductance through the point contact is proportional to the transmission probability for electrons at the Fermi energy. As the width of a point contact increases to half the de Broglie wavelength of the electrons (typically of the order of 100 nm), partial transmission through its lowest transverse mode becomes possible, and the conductance increases with decreasing partition noise. Once the lowest transverse mode in the point contact is fully transmitting, a plateau is reached in the conductance corresponding to the quantum unit of conductance (with spin degeneracy), $G_0 = 2e^2/h$, and the partition noise is completely suppressed (Fig. 3). Electrons are now steadily injected into the single mode of the point contact and are transmitted without stochasticity. Then, as the width of the point contact increases further, the higher transverse modes open up one by one with partition noise associated only with the opening mode. This is suggested by the fact that the noise normalized by the value of full shot noise corresponding to the total current shows peaks whose magnitude decreases as the transmission of the highest opening mode passes through $1/2$.

In the electron collision experiment, similar Schottky gate point contacts form the two inputs of the beam splitter, and both are set at their first conductance plateau where electrons are steadily injected with unity transmission. However, because of device non-idealities leading to finite reflection from the beam splitter, the input conductances at this plateau are less than G_0 , and the overall transmissions into the right output (where the noise is measured) are only about 35% (left input) and 30% (right input). The reflections result in additional partition noise and cause the collision noise power to increase with current. Complete noise suppression (as in the ideal model) is no longer possible because after collision, the electrons will occasionally leave through the input ports, causing the electron flux in the right output to fluctuate. These fluctuations, however, should still be smaller than the purely classical case, owing to the quantum interference.

The output noise power is plotted in Fig. 2b against the output current when each input port is biased individually and when both input ports are simultaneously biased. The linear current dependence is a sign of noise due to partitioning. Because the transmissions into the right output are about the same for the two inputs, the slopes for the individual bias cases should be and are roughly the same. On the other hand, it is evident that the slope in the case of collision is smaller. Figure 2c emphasizes the significance of this by

replotting the data as the normalized current noise relative to the expected classical collision noise determined from the sum of the individual partition noise powers. Given the transmission probabilities and using a coherent scattering formalism¹² assuming a model for the device that accounts only for the overall transmissions from inputs to outputs and disregards path length differences, the fermionic collision noise should be 52% of the classical collision noise¹³. The observed suppression of the measured collision noise to an average of 56% of the classical value therefore indicates the presence of the fermion quantum interference.

The materials and methods used in this demonstration of fermion interference should be useful for studying other fluctuation phenomena in highly degenerate electron systems where quantum statistical effects can be manifest in the second-order coherence function. Further experiments on the quantum mechanics of fermions, such as studying intensity correlations from a thermal electron source^{12,14} (analogous to the Hanbury Brown and Twiss¹⁵ experiment for photons), should be possible in this new field of quantum electron optics. □

Received 18 August; accepted 14 November 1997.

1. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics*, Vol. 3; *Quantum Mechanics* (Addison-Wesley, New York, 1965).
2. Loudon, R. in *Coherence and Quantum Optics VI* (eds Eberly, J. H. et al.) 703–708 (Plenum, New York, 1990).
3. Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
4. Zeilinger, A. *Int. Symp. on Quantum Measurements with New Experimental Techniques*, March 20–21 1997, Stanford University.
5. Liu, R. C., Odom, B., Kim, J., Yamamoto, Y. & Tarucha, S. in *23rd Int. Conf. The Physics of Semiconductors* (eds Scheffler, M. & Zimmermann, R.) 2399–2402 (World Scientific, New Jersey, 1996).
6. van Langen, S. A. & Büttiker, M. Quantum-statistical current correlations in multi-lead chaotic cavities. *Phys. Rev. B* **56**, R1680–R1683 (1997).
7. Reznikov, M., Heiblum, M., Shtrikman, H. & Mahalu, D. Temporal correlation of electrons: suppression of shot noise in a ballistic quantum point contact. *Phys. Rev. Lett.* **75**, 3340–3343 (1995).
8. van Wees, B. J. et al. Quantized conductance of point contacts in a two-dimensional electron gas. *Phys. Rev. Lett.* **60**, 848–850 (1988).
9. Wharam, D. A. et al. One-dimensional transport and the quantization of the ballistic resistance. *J. Phys. C: Solid State Phys.* **21**, L209–L214 (1988).
10. Li, Y. P., Tsui, D. C., Heremans, J. J. & Simmons, J. A. Low-frequency noise in transport through quantum point contacts. *Appl. Phys. Lett.* **57**, 774–776 (1990).
11. Kumar, A., Saminadayar, L., Glatelli, D. C., Jin, Y. & Etienne, B. Experimental test of the quantum shot noise reduction theory. *Phys. Rev. Lett.* **76**, 2778–2781 (1996).
12. Büttiker, M. Scattering theory of current and intensity noise correlations in conductors and wave guides. *Phys. Rev. B* **46**, 12485–12507 (1992).
13. Liu, R. C. *Quantum Noise in Mesoscopic Electron Transport*. Thesis, Stanford Univ., 1998.
14. Martin, Th. & Landauer, R. Wave-packet approach to noise in multichannel mesoscopic systems. *Phys. Rev. B* **45**, 1742–1755 (1992).
15. Hanbury Brown, R. & Twiss, R. Q. Correlation between photons in two coherent beams of light. *Nature* **177**, 27–29 (1956).

Acknowledgements. We thank G. Austing and T. Honda for fabrication assistance, J. Kim for help with experiments and the Joint Services Electronics Program for graduate support to R.C.L.

Correspondence should be addressed to Y.Y. (e-mail: yamamoto@loki.stanford.edu).

Picosecond discharges and stick-slip friction at a moving meniscus of mercury on glass

R. Budakian, K. Weninger, R. A. Hiller, S. J. Putterman

Physics Department, University of California, Los Angeles, California 90095, USA

At a meeting of the French Academy in 1700, Bernoulli demonstrated that swirling mercury in an evacuated flask generates light^{1,2}. He emphasized that this 'barometer light' "has not been explained since its discovery about 30 years ago" by Picard³. Here we revisit this phenomenon and find that the repetitive emission of light from mercury moving over glass is accompanied by the collective picosecond transfer of large numbers of electrons. When brought into contact with mercury, the glass acquires a net charge. This charge separation provides a force which, in our experiment in a rotating flask, drags mercury against gravity in the direction of the motion of the glass. Eventually the edge of the mercury slips relative to the glass, accompanied by a picosecond electrical discharge and a flash of light. This repetitive build-up and discharge of static electricity thus gives rise to stick-slip motion. The statistics of the intervals between events and their respective magnitudes are history-dependent and are not yet understood.

Figure 1 shows a photograph of the apparatus used for these measurements. A cylindrical glass cell is sealed and mounted so that its axis lies in the horizontal plane perpendicular to the direction of gravity. It is filled to ~10% of its volume with mercury and in this case sealed in a neon gas atmosphere at a pressure of 340 torr. In our experiments this cell is rotated about its axis at a constant speed. As seen in the photograph light is emitted all along the line of contact of the mercury and the glass on the side where the glass leaves the mercury. Rough estimates indicate a power of 20 nW per centimetre of meniscus at a speed of rotation of 1 cm s⁻¹ at the wall. This emission is easily seen with the unaided eye.

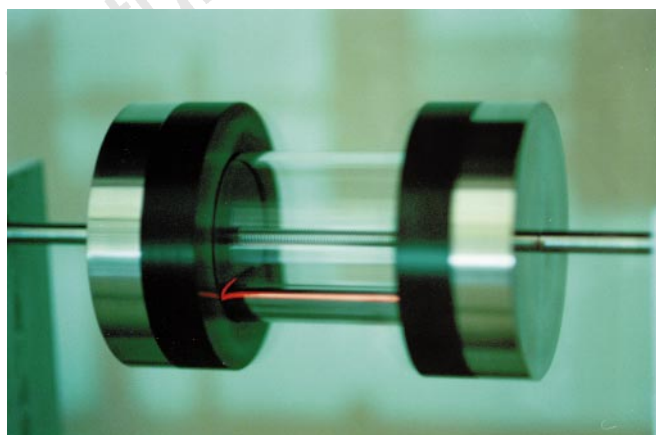


Figure 1 Photograph of the "barometer light", which is the orange line of light generated at the intersection of the mercury meniscus and the wall of the rotating glass cylinder. The rotation speed is 30° s⁻¹; the cylinder is made from Supracil with a wall thickness of 1 mm, a length of 5 cm, and an outer diameter of 2 cm. The cell is filled with 15 ml of mercury and above the mercury is neon gas at 340 torr. The light is emitted on the side where the wall leaves the mercury. With dimmed ambient lighting this effect is easily seen with the unaided eye, provided that the mercury is properly cleaned. Each black Delrin endcap is joined to the glass with an O-ring seal. We note that the Delrin-mercury interface also emits light.

Measurements were carried out using (1) a photomultiplier tube to resolve the temporal properties of the light, (2) a video camera to follow the line of contact between the mercury and glass, and (3) a capacitor (consisting of the 1-mm-diameter central conductor of an RG58 coaxial cable, with the shield cut parallel to the tip, pointing at the meniscus from the outside of the cell) to measure the charge transfer. The data shown in Fig. 2 were obtained using a nitrogen atmosphere (110 torr) above the mercury. In Fig. 2a, circles show the time dependence of the height of a 3-mm segment of the line of contact of the mercury and the glass; the electrical signals recorded by the capacitor are shown by vertical lines in this figure. Our key finding is that between electrical discharge events the mercury is moving in the direction of the glass (Fig. 2a) with a constant vertical velocity and that every discharge event, (which can have

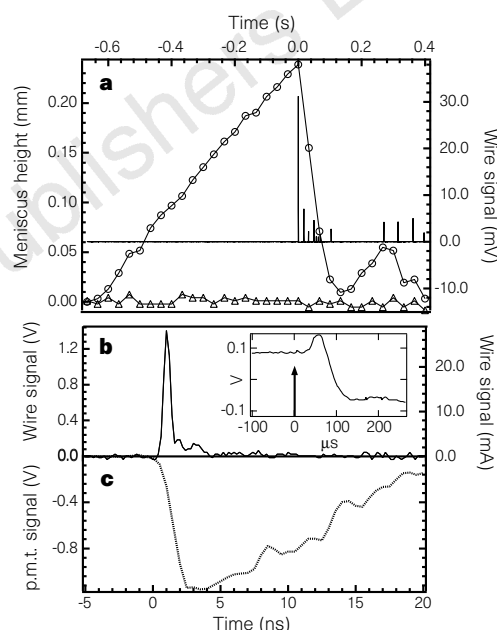


Figure 2 Correlation between stick-slip friction and picosecond electrical discharges at a glass-mercury interface. The Supracil cell is sealed with nitrogen at 110 torr and is rotated at 3° s⁻¹ by a stepper motor acting through a 10 : 1 gear reduction. The meniscus motion in a 3 mm × 3 mm region is magnified ×20 and imaged by a Cohu CCD camera. In **a**, circles indicate the meniscus height as a function of time as read off the digitized frames which are acquired at the rate of 30 pers. Also shown in **a**, keyed to the right-hand vertical axis, are the electrical discharges recorded by the capacitor probe placed near the region of the cell being imaged. The discharge from the wire is here broadened in time using a 100 KΩ terminating resistor, and digitized to memory at 4 kHz. The key conclusion to be drawn from the data in this figure is that macroscopic stick-slip events are always accompanied by collective electrical discharges and, as shown in **b**, the timescale for the electrical events is measured in picoseconds. To obtain the picosecond resolution displayed **b** the instantaneous voltage was acquired with a Tektronix 684B digital oscilloscope working at 5 × 10⁹ samples per second with a 50 Ω termination. The lower trace in **a** indicated by the triangles displays the surface motion when a deuterium lamp illuminates the entire surface with broadband ultraviolet light. In this case both electrical discharges and the stick-slip motion are suppressed. The inset in **b** shows a qualitative measure of the ~100 μs hydrodynamic timescale for meniscus motion to be completed after the discharge event (indicated by the arrow). A position-sensitive photodetector records laser light scattered off the meniscus (for this data the rotation rate is 5° s⁻¹). The timescale for light emission by the nitrogen gas that has been excited by the electrical discharge is determined by a photomultiplier tube (p.m.t.) with a rise time of 650 ps as shown in **c**. Light is collected from the mercury-glass interface using a 800-μm silica fibre which then channels it to the photodetector which is sufficiently removed so as to minimize cross-talk between its anode and the electrical discharge.

an instrument-limited rise time of 375 ps; Fig. 2b) is followed by a sudden fall in the height of the mercury. The scattering of laser light directed at the curved meniscus indicates that the hydrodynamic response of the fluid surface to the discharge-triggered slip (indicated by the arrow in Fig. 2b inset) takes $\sim 100 \mu\text{s}$ to be completed.

Our picture of the process is as follows: as portions of the glass come into contact with the mercury, electrons hop to sites on the glass where they remain as the glass separates from the mercury. This trapped static charge exerts a force on its image in the mercury which is strong enough to drag the fluid along until a sudden discharge releases this force, allowing the mercury to fall. Acquisition of the signal shown in Fig. 2b is made possible by the fact that the separated charge induces an image charge on the capacitor probe as well as in the mercury. When the discharge occurs, the capacitor discharges to ground through a 50Ω resistor, producing the voltage plotted in Fig. 2b. According to this measurement, a charge Q of $\sim 10^6$ electrons is trapped in the square millimetre being probed.

Consistent with this picture is the response of the surface of the mercury when it is exposed to strong ultraviolet illumination from a deuterium lamp. In this case the photoelectric effect prevents the charge separation required for friction between the mercury and the rotating glass wall, and the fluid remains at rest. At least for this system stick-slip friction has its origins in static electricity, a view which has in general been ruled out^{4,5}. Whether our results have a more general application to the wide range of systems that display stick-slip friction^{6,7} (and to the related phenomenon of polishing⁸) remains to be seen.

The effect of surface contaminants on interactions at a mercury-dielectric interface^{5,9–11} has been known for a long time¹². In our experiments all solid surfaces are cleaned with acetone. Dripping the mercury through acetone is sufficient to render the light-emitting state attainable. To reproducibly observe the effects reported here we allowed drops of mercury from a pipette to fall through a 25% solution of (15.8 M) nitric acid in water¹³. If the mercury-glass system is exposed to air for several days, the size of the individual

discharge events increases at first. For still older mercury the “barometer light” disappears, an effect which delayed the acceptance of Bernoulli’s demonstrations¹⁴.

The barometer light can be compared to sonoluminescence^{15–17} in that each involves the conversion of mechanical stress (rotation or sound) into light. The spectra, however, are different: the latter are broad band without lines, and the former are dominated by the emission lines of the surrounding gas^{15–18}. The phenomena are similar in that both are triggered by a collective picosecond process.

Figure 3a shows the distribution of the size of events as a function of the time between them. These data indicate that a longer time between events implies a larger size for the next event, characteristic of earthquake statistics^{19–21}. By summing all possible discharge sizes Q for a fixed time t between events, one finds that the probability per second of the next event has the form $(t/\tau^2)\exp(-t/\tau)$ where $\tau \approx 60$ ms. From the rotation rate (5°s^{-1}) and the cell diameter, this lifetime determines a characteristic length $\zeta \approx 50 \mu\text{m}$. Assuming that ζ determines the surface charge density ($\sigma \approx Q/\zeta^2$) makes the observed electrical discharge consistent with the well known equations of gas-discharge physics^{18,22–24}. That is, the resulting electric field $E \approx \sigma/\epsilon_0$ where ϵ_0 is the electrical permittivity of vacuum, is sufficient to accelerate an electron to the ionization energy χ of a gas atom, typically 10 eV, in a mean free path l (that is, $eEl \approx \chi$, where e is the charge on an electron, $l \approx (p_0/p) \mu\text{m}$, p_0 is 1 atm, and p is the gas pressure in the cell). The collisions between the accelerated electrons and the surrounding gas leave the atoms/molecules in excited states. Light-emitting transitions from these states have lifetimes ranging well into nanoseconds, as shown in Fig. 2c. Although the existence of the discharge is consistent with established theory we propose that the following processes remain unexplained: (1) the mechanism of charge transfer from mercury to the glass^{25,26} (tunnelling has been suggested^{27,28}); (2) the distribution of the excess charge on the glass (that is, the characteristic length ζ) which accounts for the ‘static cling’, discharge and slip events; and (3) the picosecond quenching of the charge separation.

In Fig. 3b we show the distribution of events for the case where the fluid is illuminated with ultraviolet light that is strong enough to affect the motion but is not as intense as was used to suppress friction as shown in Fig. 2a. In this case the illumination can be adjusted so that the events are markovian. The intrinsic process that controls the distribution gives in Fig. 3b has the form $(t^2/2\tau^3)\exp(-t/\tau)$ with a lifetime τ of 3 ms (so without ultraviolet illumination there are fewer but larger events). The distribution of the charge amplitude of intrinsic events (obtained by plotting as a histogram all values of Q shown in Fig. 3b) quickly rises to a maximum at $Q_0 \approx 10^5 e$ and then decays as $\exp(-2Q/Q_0)$. Finally, as part of our effort to control this phenomenon, we have vibrated in the vertical direction a glass capillary which has one end immersed in a pool of mercury. In the centre of the capillary is a wire which generates signals as high as 30 V (across 50Ω) with picosecond rise times that are again instrument-limited. Shown in Fig. 3c is the distribution of events for this means of using triboelectricity to generate a string of picosecond electrical pulses.

Contact electrification in this cell is also affected by the gas pressure. When reduced to the range of 30 mtorr the light emission becomes diffuse and more characteristic of a corona discharge²⁹ and in such an atmosphere the stick-slip friction is eliminated. The nature of friction between glass and mercury at lower pressures (say 5 mtorr)—where well controlled experiments²⁸ have shown that large charge separations can build up as a result of the contact between surfaces—is beyond the capabilities of our current experiment.

Our measurements show that for mercury sliding on glass, stick-slip friction is due to the build-up of electric charge and its subsequent neutralization on a picosecond timescale. Small fluid velocities can generate a continuous stream of events (20 cyclings of charge per second at 2 minutes per revolution). This may be an ideal

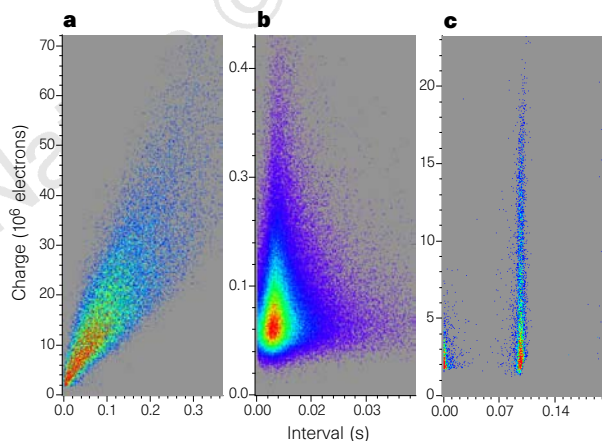


Figure 3 Plot of discharge events in terms of the time to the next event and its strength. The distribution of discharge events is obtained for: **a**, a Supracil quartz cell sealed with nitrogen at 110 torr rotating at 5°s^{-1} ; and **b**, the same arrangement (rotating at 2°s^{-1}) but subjected to illumination with ultraviolet light. The intensity of illumination is substantially less than was used to suppress stick-slip friction in Fig. 2a. In **c** is displayed the distribution for events generated by a 1-cm-long glass capillary of 2-mm outer diameter fused over platinum wire which oscillates vertically at 10 Hz in a pool of mercury in a 200 torr neon atmosphere. In **a** and **b** the signal across $1 \text{ M}\Omega$ is amplified and low-pass-filtered before being digitized. The noise level after filtering the 60 Hz component is one-seventh of the trigger level. Data is acquired for ~ 1 h and the rate of (the smaller) events in **b** is over ten times greater than in **a**. In **c** is displayed the result of 7,200 events. In all panels, the density of points is highest for red and lowest for purple.

system to study contact electrification and the formative time of sparks^{22–24}. We propose that picosecond frictional-electricity may yield a useful characteristic of surfaces. One can wonder if liquid helium rubbing against caesium would display these effects. □

Received 25 March; accepted 7 October 1997.

- Bernoulli, J. I. Sur le phosphore du baromètre. *Histoire Acad. Roy. Paris* 5–8 (1700); 1–8 (1701).
- Harvey, E. N. *A History of Luminescence* (Am. Philosophical Soc., Philadelphia, 1957).
- Picard, J. Sur la lumière du baromètre. *Mem. Acad. Roy. Sci.* 2, 202–203 (1676).
- Krim, J. Friction at the atomic scale. *Sci. Am.* 275, 48–51, 54–56 (1996).
- Harper, W. R. *Contact and Frictional Electrification* (Clarendon, Oxford, 1967).
- Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* (Clarendon, Oxford, 1986).
- Bhushan, B., Israelachvili, J. N. & Landman, U. Nanotribology: friction, wear and lubrication at the atomic scale. *Nature* 374, 607–616 (1995).
- Rabinowicz, E. *Friction and Wear of Materials* (Wiley, New York, 1965); Polishing. *Sci. Am.* 218, 91–99 (1968).
- Kemball, C. The adsorption of vapours on mercury. IV. Surface potentials and chemisorption. *Proc. R. Soc. Lond. A* 201, 377–391 (1950).
- Hays, D. A. in *Conference Series No. 48* 265–272 (Conf. Ser. No. 48, Inst. Phys., 1979).
- Dybwad, G. L. & Mandeville, C. E. Generation of light by the relative motion of contiguous surfaces of mercury and glass. *Phys. Rev.* 161, 527–532 (1967).
- Rayleigh, Lord Experiments upon surface-films. *Phil. Mag.* 33, 363–373 (1892).
- Bhushan, B., Israelachvili, J. N. & Landman, U. Nanotribology: friction, wear and lubrication at the atomic scale. *Nature* 374, 607–616 (1995).
- Bernoulli, J. I. Nouvelle maniere de rendre les baromètres lumineux. *Mem. Acad. Roy. Paris* 178–190 (1700).
- Barber, B. P. & Putterman, S. J. Observation of synchronous picosecond sonoluminescence. *Nature* 352, 318–320 (1991).
- Hiller, R. A., Putterman, S. J. & Barber, B. P. Spectrum of synchronous picosecond sonoluminescence. *Phys. Rev. Lett.* 69, 1182–1184 (1992).
- Barber, B. P. *et al.* Defining the unknowns of sonoluminescence. *Phys. Rep.* 281, 65–143 (1997).
- Meek, J. M. & Craggs, J. D. (eds) *Electrical Breakdown of Gases* (Wiley, New York, 1978).
- Carlson, J. M., Langer, J. S. & Shaw, B. E. Dynamics of earthquake faults. *Rev. Mod. Phys.* 66, 657–670 (1994).
- Demirel, A. L. & Granick, S. Friction fluctuations and friction memory in stick-slip motion. *Phys. Rev. Lett.* 77, 4330–4333 (1996).
- Burridge, R. & Knopoff, L. Model and theoretical seismicity. *Bull. Seismol. Soc. Am.* 57, 341–371 (1967).
- Rees, J. A. *Electrical Breakdown of Gases* (Macmillan, New York, 1973).
- Penning, F. M. *Electrical Discharges in Gases* (Cleaver Hume, London, 1957).
- Loeb, L. B. *Fundamental Processes of Electrical Breakdown in Gases* (Wiley, New York, 1939).
- Moore, A. D. *Electrostatics* (Doubleday, New York, 1968).
- Terris, B. D., Stern, J. E., Ruger, D. & Mamin, H. J. Contact electrification using force microscopy. *Phys. Rev. Lett.* 63, 2669–2672 (1989).
- Lowell, J. Tunneling between metals and insulators and its role in contact electrification. *J. Phys. D* 12, 1541–1554 (1979).
- Kwetkus, B. A., Sattler, K. & Siegmann, H.-C. Gas breakdown in contact electrification. *J. Phys. D* 25, 139–146 (1992).
- Raizer, Y. P. *Gas Discharge Physics* (Springer, Berlin, 1991).

Acknowledgements. We thank L. Knopoff, T. Erber, J. Raffelski, G. Morales, S. Cowley, R. Löfstedt and P. H. Roberts for discussions, and E. Adams and C. Hiller for archival assistance. This work was supported by the US NSF and the US Department of Energy.

Correspondence should be addressed to S.J.P.

The structure of a new phase of ice

C. Lobban*, J. L. Finney* & W. F. Kuhs†

* Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

† MKI, Universität Göttingen, Goldschmidtstrasse 1, 37077 Göttingen, Germany

Ice has eleven known crystalline phases (Fig. 1), in which the water molecules are linked through hydrogen bonds into tetrahedral frameworks¹. This uncommonly large number of different solid phases attests to the structural versatility of the water molecule. Here we report the identification of a new, twelfth phase of crystalline ice in the pressure range 0.2–0.6 GPa. The topology of this phase is unlike that of any of the known phases, and contains a mixture of five- and seven-membered rings of water molecules. It has a density similar to that of ice IV, which also occurs in this pressure range within the stability region of ice V. Both phases are likely to be metastable with respect to the less-dense ice V. This region of the water phase diagram thus provides a potential model system for experimental and theoretical studies of metastability.

As the water molecule frameworks in ice structures are compressed under pressure, the strain in the hydrogen bonds results in

bond breakage, with subsequent water molecule rearrangements forming a different structure of higher density. At pressures of less than 0.2 GPa, the water molecules adopt very open structures (ices I and II), whereas at pressures above 0.6 GPa, the ice structures accommodate the increased external pressure by forming two interpenetrating frameworks (ices VI, VII and VIII). In the interesting intermediate region, the increased pressure is accommodated through strained single frameworks, with the strains relieved in different structural ways depending on the precise pressure and temperature conditions.

Ice V is believed to be thermodynamically stable between 0.35 and 0.60 GPa, and below ~265 K (Fig. 1). It adopts a complicated structure with hydrogen bonds being substantially bent². A 'metastable' form of ice, ice IV, which exists within the stability region of ice V, was first proposed by Bridgman³. This, together with the high strain in the ice V structure, suggests that other ice structures might be almost as stable under similar conditions of temperature and pressure. Engelhardt and Whalley, in establishing the stability region of ice IV, found evidence for the existence of other metastable phases⁴. Unfortunately, these phases proved very difficult to form, and could not be retained long enough for useful study. Similarly, in experiments in this pressure region primarily aimed at ices III, IV and V⁵, we have observed neutron diffraction features that we have been unable to identify as any known ice or clathrate phase.

In following up one such neutron diffraction experiment, we have formed, retained and positively identified a new phase of ice within the stability region of ice V. The phase remained stable long enough to allow data to be collected of sufficient quality for detailed structure analysis.

Roughly 1.5 ml of D₂O (99.9% D) was loaded into a (TiZr) gas pressure cell, previously filled loosely with 0.1–0.2 g of silica wool to ensure a good powder on freezing. Argon gas was used to increase the pressure to 0.55 GPa, and the cell lowered into a cryostat located on a diffractometer. The sample was cooled quickly to 270 K, and then at a rate of 2.5 K hr⁻¹ to 260 K. This rate of cooling seems to be important for nucleation of the new phase; a slightly faster rate of 5 K hr⁻¹ resulted in the formation of ice V. At 260 K, the pressure was topped up to 0.55 GPa. After periods of between 1 and 3 hours at this condition, the water froze to the new ice phase. On no occasion has the preparation of the new phase been unsuccessful when following this recipe. As we have found when preparing other phases of ice using argon as the pressurizing medium, a small amount of argon clathrate was also formed⁵. The degree of clathrate in the liquid phase was small, and is not thought to be important in the nucleation and growth of the new phase. The clathrate was formed predominantly from the solid phase after the new phase had been formed.

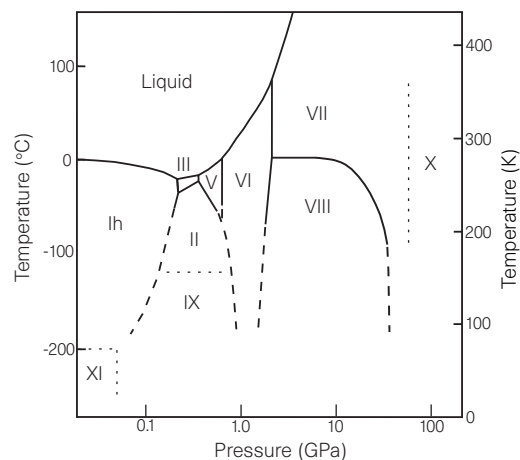


Figure 1 The phase diagram of ice.

Neutron powder data were collected at different times on different samples on three neutron powder diffractometers: PEARL and HRPD at ISIS, and D2B at ILL. The structure of the new phase was solved from the primary data set collected on PEARL. For the data collected in the range 1 to 4 Å, a tetragonal unit cell, $a = 8.304$ Å, $c = 4.024$ Å, was found to be the most promising indexing scheme. The highest symmetry space group that meets the reflection conditions is $I\bar{4}2d$. The density was estimated in two ways. Once data had been collected on the new phase, the sample was cooled to induce the formation of ice V. The transition to ice V resulted in a

pressure increase arising from an increase in sample volume, implying ice V to be less dense. In addition, the density was also estimated from abrupt drops in pressure associated with the sample freezing. The drop recorded for the new phase was similar to that found for ice IV and larger than that for ice V. From this, an estimated density led to a figure of 12 water molecules per unit cell which was permitted under the space group $I\bar{4}2d$ with the water molecules occupying special positions. The water molecules were then positioned to make physical sense in terms of bond lengths and angles and the model refined using the GSAS system⁶ (Fig. 2).

Table 1 Fractional coordinates and thermal factors for the water molecules found in the new ice structure

	<i>x</i>	<i>y</i>	<i>z</i>	<i>B</i> ₁₁	<i>B</i> ₂₂	<i>B</i> ₃₃	<i>B</i> ₁₂	<i>B</i> ₁₃	<i>B</i> ₂₃
O(1)	0	0	0	3.52(24)	3.52(24)	2.66(40)	0	0	0
O(2)	0.3643(4)	0.25	0.125	4.33(22)	1.85(17)	1.67(14)	0	0	−0.84(10)
D(3)	0.0419(6)	0.0871(4)	−0.1340(13)	3.02(19)	4.58(38)	3.42(28)	0.99(18)	−1.85(20)	−0.66(19)
D(4)	0.2922(5)	0.2181(4)	0.3090(13)	4.73(35)	3.66(20)	1.99(32)	−0.09(15)	−0.22(12)	−0.20(11)
D(5)	0.4202(6)	0.3327(5)	0.2226(15)	3.55(22)	4.54(22)	2.79(27)	1.23(17)	−0.30(15)	−1.50(15)

Table 2 Intra- and intermolecular distances, angles and multiplicities for the new ice phase at 0.50 GPa and 260 K

Bond	Distance (Å)	Multiplicity	Bond	Distance (Å)	Multiplicity
O(1)–D(3)	0.967(4)	4	O(1)···O(2)	2.803(1)	6
O(2)–D(4)	0.988(4)	2	O(2)···O(2)	2.766(4)	2
O(2)–D(5)	0.918(5)	2			
Bond angle	Angle (°)	Multiplicity	Bond angle	Angle (°)	Multiplicity
D(3)–O(1)–D(3)	108.1(3)	4	O(2)···O(1)···O(2)	106.85(2)	4
D(3)–O(1)–D(3)	112.2(6)	2	O(2)···O(1)···O(2)	114.86(3)	2
D(4)–O(2)–D(4)	105.4(8)	1	O(2)···O(2)···O(2)	93.34(2)	1
D(4)–O(2)–D(5)	115.3(3)	2	O(2)···O(2)···O(1)	131.87(5)	2
D(4)–O(2)–D(5)	100.7(5)	2	O(2)···O(2)···O(1)	83.36(2)	2
D(5)–O(2)–D(5)	119.2(7)	1	O(1)···O(2)···O(1)	132.58(1)	1

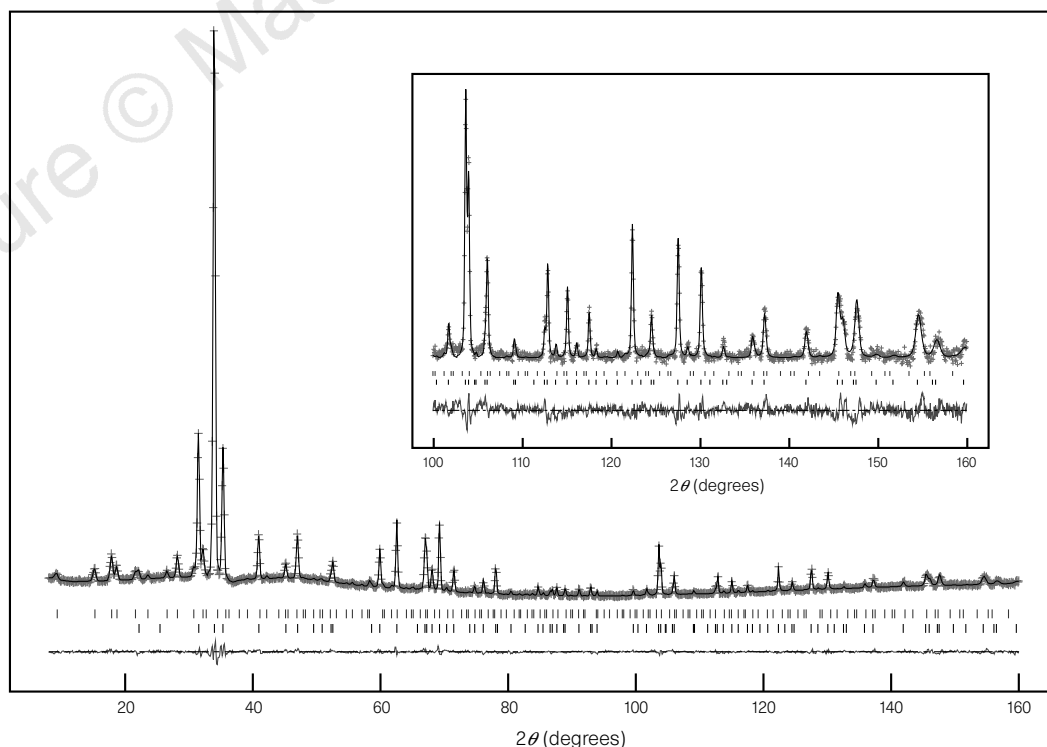


Figure 2 Observed, calculated and difference profiles for the new ice phase at 0.50 GPa and 260 K. The structure was refined free of any restraints, with full deuteron disorder, using the GSAS system⁶. As the low *d*-spacing data of the PEARL data set gave rise to an unphysical temperature factor for one deuteron, the final refinement was performed on the data collected on D2B ($\lambda = 1.594$ Å). Final residuals were $R_{wp} = 0.0139$, $R_p = 0.0098$. The upper plot represents the

observed data (crosses) with the calculated fit (solid line) overlying them. The lower plot represents the difference between observed and calculated intensities, with the tick marks between plots the reflection positions of the argon clathrate (upper) and new ice phase (lower). The amount of argon clathrate in the sample as determined by the refinement is 14%.

Table 3 Density and bond bending found for ices IV, V and the new phase at 0.50 GPa and 260 K

Phase	ρ (g cm ⁻³)	$\langle(\delta\theta)^2\rangle^{1/2}$
New phase	1.4365(1)	18.868
Ice IV	1.4361(1)	15.058
Ice V	1.4021(1)	19.067

$\langle(\delta\theta)^2\rangle^{1/2}$ is the r.m.s. of O...O...O bond angle deviations from the ideal tetrahedral angle of 109.5°, and ρ is density.

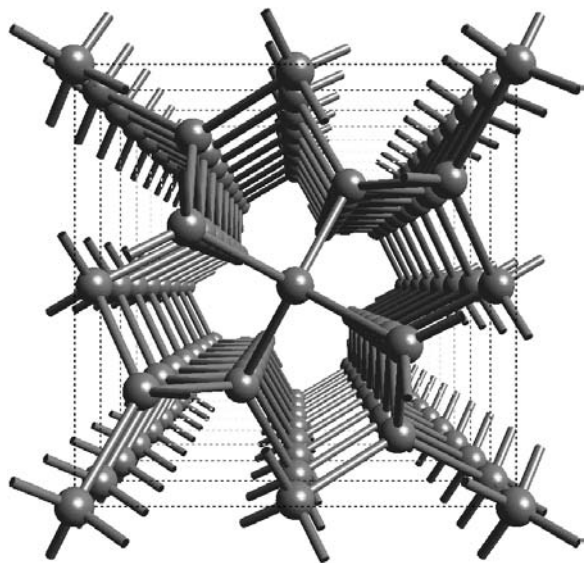


Figure 3 The oxygen framework of the new ice structure viewed down the *c*-axis.

The new phase is tetragonal, $a = 8.304 \text{ \AA}$, $c = 4.024 \text{ \AA}$, space group $I4_2d$. The unit cell comprises 12 water molecules, arranged to form a single-network tetrahedral structure. Fractional coordinates and refined bond lengths and angles are given in Tables 1 and 2 and the structure is shown in Fig. 3. The bond lengths are of the order expected for an ice in which the hydrogen atoms are positionally disordered: diffraction data give spatially averaged atomic positions which may differ from true local positions depending on the details of the disorder. Water molecules of type O(2) form zigzag chains which run parallel to the *c*-axis, and oscillate roughly parallel to either the *a*- or *b*-axis. These chains are then linked together by water molecules of type O(1), resulting in channels, roughly pentagonal in shape, running parallel to the *c*-axis. This configuration results in a density of 1.437 g cm^{-3} , which is greater than that of ice V at 1.402 g cm^{-3} , yet comparable with that of ice IV, 1.436 g cm^{-3} . For ice IV, the increase in density over ice V is achieved through a form of interpenetration, in which a hydrogen bond passes through a hexagonal ring of water molecules. For the new ice phase, no interpenetration of bonds is observed. Instead, the increase in density over ice V is achieved through additional bending of the hydrogen bonds (Table 3).

There can be no suggestion that this new phase is ice IV. Although the structure of ice IV in the literature was determined from samples recovered at low temperature to ambient pressure⁷, in other measurements we have also formed and refined the structure of ice IV under pressure⁵, and can thus confirm that the structure of ice IV under these pressure and temperature conditions is that proposed by Engelhardt and Kamb⁷.

The designation of phases I through to XI now seems to be accepted. Although the label of ice XII has already been used by Sirota and Bizhigitov⁸ for a low-temperature phase in the pressure range of ice VI and VIII, the evidence for this phase is from relative

volume changes observed in piston-in-cylinder cell measurements, with no direct structural technique having been used in its investigation. Despite the efforts of several groups, the phase suggested by Sirota and Bizhigitov⁸ has never been confirmed. We therefore provisionally propose that this new phase reported here be named ice XII. (It seems sensible to follow the generally accepted view that Roman numerals be used only for experimentally established crystalline phases, with crystallographic or at least spectroscopic evidence required to fulfil the criterion of 'experimental evidence'.)

The new picture that emerges is of two disordered phases of ice, ices IV and XII, that are very similar in density yet are topologically very different. Both phases are likely to be metastable with respect to ice V, but have somewhat higher density. The existence of 5- and 7-fold rings is interesting, although not unexpected: such local geometries are entirely consistent with the roughly tetrahedral geometry of the intermolecular interaction, and are found in some other ice structures (for example, ices III and V). The enhanced richness of the ice phase diagram in the medium pressure range is another demonstration of the versatility of the water molecule in building a variety of fully hydrogen-bonded structures. The seemingly delicate balance of enthalpic and entropic contributions to the total energy will allow us to test critically the viability of water potential functions used widely in computer simulations of chemical and biomolecular systems. The water system also seems to be an excellent candidate for general studies of metastability, including both thermodynamic and kinetic aspects of phase formation. □

Received 26 August; accepted 27 October 1997.

1. Hobbs, P. V. *Ice Physics* (Clarendon, Oxford, 1974).
2. Kamb, B., Prakash, A. & Knobler, C. Structure of ice V. *Acta Cryst.* **22**, 706–715 (1967).
3. Bridgman, P. W. The *P*–*V*–*T* relations of the liquid, and the phase diagram of heavy water. *J. Chem. Phys.* **3**, 597–605 (1935).
4. Engelhardt, H. & Whalley, E. Ice IV. *J. Chem. Phys.* **56**, 2678–2684 (1972).
5. Lobban, C. *Neutron Diffraction Studies of Ices*. Thesis, Univ. London (1998).
6. Larsen, A. C. & Von Dreele, R. B. *GSAS—General Structure Analysis System*. (University of California, 1985).
7. Engelhardt, H. & Kamb, B. Structure of ice IV, a metastable high-pressure phase. *J. Chem. Phys.* **75**, 5887–5899 (1981).
8. Bizhigitov, T. B. & Sirota, N. N. Low-temperature modification of high-pressure ice. *JETP Lett.* **44**, 417–419 (1986).

Acknowledgements. We thank CCLRC and the Institut Laue-Langevin for access to the ISIS and ILL neutron facilities respectively, including generous allocation of the director's discretionary time which was crucial in obtaining data of quality adequate for the structure refinement. We thank also J. Dreyer and T. Cooper (ISIS) and L. Melesi (ILL) for their extensive efforts on the high-pressure equipment.

Correspondence and requests for materials should be addressed to either J.L.F. (e-mail: j.finnay@ucl.ac.uk) or W.F.K. (e-mail: kuhse@silly.uni-mki.gwdg.de).

Silicate regulation of new production in the equatorial Pacific upwelling

Richard C. Dugdale & Frances P. Wilkerson

Romberg Tiburon Centers, San Francisco State University, PO Box 855, Tiburon, California 94920, USA

Surface waters of the eastern equatorial Pacific Ocean present the enigma of apparently high plant-nutrient concentrations but low phytoplankton biomass and productivity¹. One explanation for this 'high-nitrate, low-chlorophyll' (HNLC) phenomenon has been that growth is limited by iron availability^{2,3}. Here we use field data and a simple silicon-cycle model⁴ to investigate the HNLC condition for the upwelling zone of this ocean region. Measured silicate concentrations in surface waters are low and largely invariant with time, and set the upper limit on the total possible biological utilization of dissolved inorganic carbon.

Chemical and biological data from surface waters indicate that diatoms—silica-shelled phytoplankton—carry out all the ‘new production’ (nitrate uptake)⁵. Smaller phytoplankton (picoplankton) accomplish most of the total primary production, largely fuelled by nitrogen regenerated in reduced forms as a result of grazing by zooplankton. The model predicts values of new and export production (the production exported to below the euphotic zone) that compare well with measured values⁶. New and export production are in balance for biogenic silica, whereas new production exceeds export for nitrogen. The HNLC condition in the upwelling zone can therefore be understood to be due to a chemostat-like regulation of nitrate uptake by upwelled silicate supply to diatoms: ‘low-silicate HNLC’. These results are not inconsistent with observations of iron-fertilized diatom growth during *in situ* experiments in ‘low-iron HNLC’ waters outside this upwelling zone^{2,3}, but reflect the role of different supply rates of iron and silicate in determining the nature of the HNLC condition.

The equatorial upwelling zone (EUZ) of the eastern equatorial Pacific Ocean—where biological drawdown of CO₂ is low due to weak primary production—is a net source of CO₂ to the atmosphere. The EUZ is characterized by relatively continuous upwelling driven by the trade winds; this has led to it being considered as a chemostat culture system⁷, in which nutrient supply regulates phytoplankton growth. Modelling and field data show⁴ that open-ocean upwelling systems tend to become silicate-limited through

differential export of Si compared to N. Zooplankton grazing on diatoms leads to much greater regeneration of N than silicate (Si(OH)₄²⁻) in euphotic zone waters, so that the disappearance ratios of NO₃:Si(OH)₄²⁻ are <1 (as low as 0.25 in Southern Ocean surface waters⁸). These systems are eventually driven into silicate limitation by the rate of supply of silicate. Si(OH)₄²⁻ stabilizes at a regulating concentration below that of NO₃ through a feedback process that sets up the steady-state balance between nutrient supply rate and population size (nutrient demand).

The ratio of Si(OH)₄²⁻:NO₃ in equatorial upwelled source waters is usually less than the 1:1 ratio⁹ realized by diatoms, and studies have shown¹⁰ Si(OH)₄²⁻ to be transported into the equatorial upwelling zone at a rate of 0.8 times that of NO₃. The special role of diatom production (Si(OH)₄²⁻ drawdown) in the equatorial drawdown of total CO₂ (ΣCO₂) can be seen from the similarity in slopes of equatorial ΣCO₂ versus Si(OH)₄²⁻ and NO₃ (Fig. 1a, b). Using the approach of Broecker and Peng¹¹, these regressions show this relationship to be true for both deeper samples, which tend to have a slope of near Redfield (C/N = 6.6), and for near-surface data where ΣCO₂ has been reduced by exchange with the atmosphere¹¹. The higher intercepts in Fig. 1b show that the maximum ΣCO₂ drawdown due to biological production is a function of the ΣCO₂ versus Si(OH)₄²⁻ rather than NO₃ relationship; NO₃ uptake appears to cease when Si(OH)₄²⁻ falls to some low concentration (~2 mmol m⁻³; Fig. 1c). The NO₃ versus Si(OH)₄²⁻ slope for the

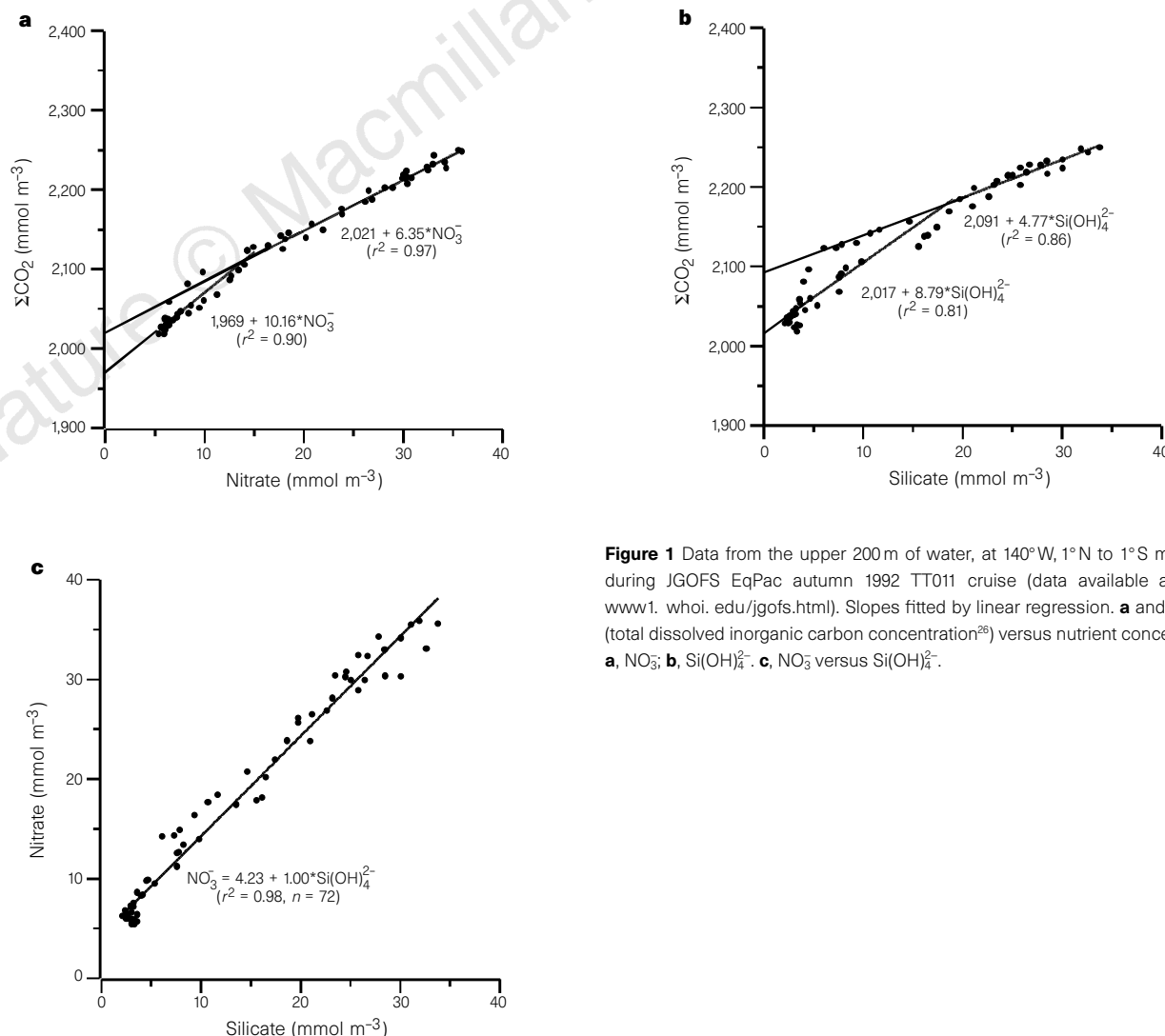


Figure 1 Data from the upper 200 m of water, at 140°W, 1°N to 1°S measured during JGOFS EqPac autumn 1992 TT011 cruise (data available at <http://www1.who.edu/jgofts.html>). Slopes fitted by linear regression. **a** and **b**, ΣCO₂ (total dissolved inorganic carbon concentration²⁶) versus nutrient concentration; **a**, NO₃⁻; **b**, Si(OH)₄²⁻. **c**, NO₃⁻ versus Si(OH)₄²⁻.

The silicate pump model⁴ was initiated for the EUZ after adding a picoplankton loop (Fig. 2) with no NH_4^+ recycled to the diatoms. Si dissolution rates were assumed to be negligible as there are

	mmol N m ⁻² d ⁻¹	mmol Si m ⁻² d ⁻¹	mg C m ⁻² d ⁻¹
Inputs			
Upwelling supply (NO ₃ ⁻ , Si(OH) ₄ ⁰)	3.97	2.36	
Diatom uptake rate (NO ₃ ⁻ , Si(OH) ₄ ⁰)	2.36	2.36	
Outputs			
Grazing rate	2.29	2.29	
Faecal pellet production rate	0.69	2.29	
Sinking rate	0.034	0.034	
Upwelling dilution rate	0.034	0.034	
Export production*	0.76	2.36	
Predation rate	0.69	0	
New production	2.36	2.36	186.91
(that is diatom production)			
Regenerated production	9.20	0	728.64
(that is, picoplankton production)			
NH ₄ ⁺ supplied from grazing	0.92		
on diatoms			
NH ₄ supplied from micrograzers	8.28		
Total production†	11.56		915.55

† Calculated as new production and regenerated production.

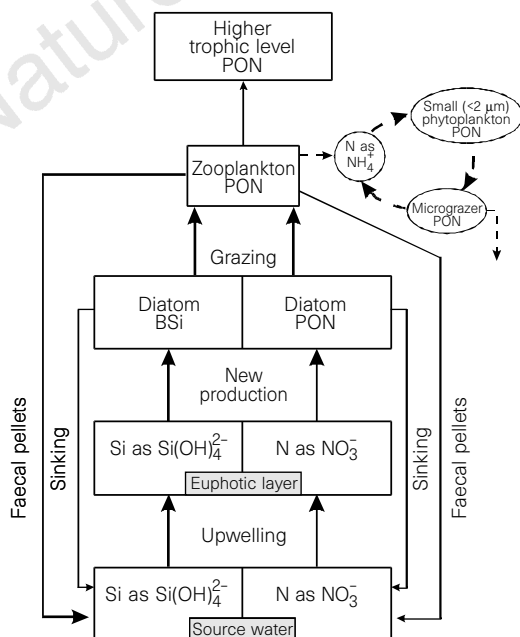


Figure 2 Silicate pump model⁴ modified to include a picoplankton/micrograzer loop. PON, particulate organic nitrogen.

Euphotic-layer $\text{Si}(\text{OH})_4^{2-}$ concentrations measured during three cruises (WEC 88, EqPac spring 1992 and EqPac autumn 1992) to the EUZ showed little variability ($2.65\text{--}2.77\text{ mmol m}^{-3}$) compared to the higher and more variable NO_3^- concentrations ($2.76\text{--}5.93\text{ mmol m}^{-3}$). These observations are consistent with chemostat conditions, in which the relative concentrations of the required nutrients in the feed medium determine which nutrient will be limiting (silicate in this case), and the loss rates set the growth rate of the organism. Regulation of the diatom growth rate, and consequently of the new production system, requires that the growth rate and the limiting substrate concentration define a point on an operating curve (typically a Michaelis–Menten relationship). To identify the operating point for EUZ an estimate of the specific silicate uptake ($V_{\text{Si}(\text{OH})_4^{2-}} = 0.82\text{ d}^{-1}$) was obtained using ^{15}N data¹⁴, assuming nitrate uptake ($\rho_{\text{NO}_3^-}$, $\text{mmol l}^{-1}\text{ d}^{-1}$) = silicate uptake ($\rho_{\text{Si}(\text{OH})_4^{2-}}$, $\text{mmol l}^{-1}\text{ d}^{-1}$) and dividing $\rho_{\text{Si}(\text{OH})_4^{2-}}$ by BSi (0.043 mmol m^{-3}). The steady-state silicate concentration for this value of $V_{\text{Si}(\text{OH})_4^{2-}}$ calculated with a half-saturation constant $K_s = 3\text{ mmol m}^{-3}$ and $V_{\text{max}} = 1.92\text{ d}^{-1}$ is 2.24 mmol m^{-3} , in good agreement with the silicate range observed above ($2.65\text{--}2.77\text{ mmol m}^{-3}$). The $V_{\text{Si}(\text{OH})_4^{2-}}$ value compares well with the directly measured diatom growth rate¹⁹ ($\mu = 0.67\text{ d}^{-1}$) and both fall close to the Michaelis–Menten curve. The diatom population is regulating at about half the maximal uptake and with a substrate concentration near K_s . The ability of the diatom population to achieve steady-state growth rates of $\sim 1\text{ d}^{-1}$, sufficient to offset the sum of the loss rates, shows that limitation by other factors such as iron need not be invoked. Oceanic diatoms can reach maximal growth rates of 1 d^{-1} with low iron²³. The supply of new iron in the upwelling water²² is

apparently sufficient to allow diatom growth at the levels required for steady state.

Additions of iron in the EUZ may change the operating point of the silicate uptake curve, for example, by increasing the initial slope. The transient effect would be a small net increase in diatom growth rate over grazing rate or increase in biomass, as has been observed²⁴ after 4–5 days. A longer-term effect of this change in operating point would be to decrease the euphotic-layer silicate concentration, resulting in an increase of the source to euphotic-zone silicate gradient, and an increase in new production. This gradient increase is always relatively small and therefore any increase in new production minimal, because the concentration of euphotic-layer silicate in the EUZ is already close to or at K_s concentrations. Although our results appear to conflict with those of the IronEx open-ocean experiments in which clear responses to iron additions were observed^{2,3}, the contradictions can be resolved by recognizing the important differences in conditions of JGOFS EqPac and IronEx studies. In the EUZ, new iron is fed into the surface waters, along with new $\text{Si}(\text{OH})_4^{2-}$ and NO_3^- fuelling the diatom productivity observed. The IronEx experiments were conducted outside the EUZ in the South Equatorial Current, where the atmosphere is the only source of new iron and the area is prone to chronic iron limitation. Although $\text{Si}(\text{OH})_4^{2-}$ values were not reported for IronEx, they would be lower than NO_3^- , as the source water for this current arises from the low-silicate HNLC region offshore from Peru⁴. Because the response of NO_3^- uptake in IronEx II was seen exclusively in the larger ($>5\ \mu\text{m}$) algal size-class (diatoms)³, our results suggest that even with iron enrichment, the maximum drawdown of CO_2 in this region would eventually be limited by silicate availability.

To increase CO_2 drawdown in the EUZ and reach coastal upwelling values of new production, an order-of-magnitude increase in upwelling rate combined with an increase in source silicate concentration would be necessary⁴. The ^{15}N measured new production¹⁴ during the EqPac autumn 1992 study was $0.03\ \text{mmol N m}^{-3}\ \text{d}^{-1}$, about two orders of magnitude below the range of new production for coastal upwelling²⁵, $0.4\text{--}4.2\ \text{mmol N m}^{-3}\ \text{d}^{-1}$. As the euphotic-zone silicate concentration is constant (a result of silicate regulation), the source concentration of silicate alone determines the vertical gradient of this species if no changes in mixed-layer depth occur. Under these conditions, the source silicate concentration and the upwelling rate are the remaining variables determining the new production rate. By elevating both source silicate concentration (for example, from 5.67 to $20\ \text{mmol m}^{-3}$) and the upwelling rate, w (from 1 to $10\ \text{m d}^{-1}$), the new production would reach $1.73\ \text{mmol m}^{-3}\ \text{d}^{-1}$ (coastal upwelling levels²⁵). Neither variable alone can achieve such rates (that is, $\text{Si}(\text{OH})_4^{2-} = 20\ \text{mmol m}^{-3}$, $w = 1\ \text{m d}^{-1}$ gives $\rho_{\text{NO}_3^-} = 0.17\ \text{mmol m}^{-3}\ \text{d}^{-1}$, and increasing w to $10\ \text{m d}^{-1}$ with $\text{Si}(\text{OH})_4^{2-} = 5.67\ \text{mmol m}^{-3}\ \text{d}^{-1}$ gives $\rho_{\text{NO}_3^-} = 0.3\ \text{mmol m}^{-3}\ \text{d}^{-1}$).

The EUZ has characteristics that are operationally analogous to chemostat culture systems, specifically (1) constant limiting nutrient concentrations with variable non-limiting concentrations, (2) growth rates as a function of loss rates; and (3) an operating point on a known nutrient/growth-rate curve. The “excess NO_3^- ” previously observed¹ that led to the classification of the eastern equatorial Pacific as HNLC can now be understood as the consequence of silicate limitation, because new N is taken up only by diatoms. Potential new production based on nitrate concentration is not realized. This area, like some other HNLC areas described previously⁴, should be classified as low-silicate HNLC to avoid the impression that all ‘major nutrient’ concentrations are sufficiently high to support new production. The results obtained in the IronEx locations of the eastern equatorial Pacific should not be extrapolated to the EUZ, which has a source of new iron and in which the potential for increased new and export production, by whatever means, is very small. □

Received 28 February; accepted 21 October 1997.

- Barber, R. T. & Chavez, F. P. Regulation of primary productivity rate in the equatorial Pacific. *Limnol. Oceanogr.* **36**, 1803–1815 (1991).
- Martin, J. H. *et al.* Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- Dugdale, R. C., Wilkerson, F. P. & Minas, H. J. The role of a silicate pump in driving new production. *Deep-Sea Res.* **42**, 697–719 (1995).
- Dugdale, R. C. & Goering, J. J. Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnol. Oceanogr.* **12**, 196–207 (1967).
- Murray, J. W., Johnson, E. & Garside, C. A.U.S. JGOFS Process Study in the equatorial Pacific (EqPac): Introduction. *Deep-Sea Res.* **42**, 275–293 (1995).
- Frost, B. W. & Franzen, N. C. Grazing and iron limitation in the control of phytoplankton stock and nutrient concentration: a chemostat analogue of the Pacific equatorial upwelling zone. *Mar. Ecol. Prog. Ser.* **83**, 291–303 (1992).
- Minas, H. J. & Minas, M. Net community production in “High Nutrient-Low Chlorophyll” waters of the tropical and Antarctic Oceans: grazing versus iron hypothesis. *Oceanol. Acta* **15**, 145–162 (1992).
- Wilkerson, F. P. & Dugdale, R. C. Silicate versus nitrate limitation in the equatorial Pacific estimated from satellite-derived sea-surface temperatures. *Adv. Space Res.* **18**, 81–89 (1996).
- Ku, T.-L., Luo, S., Kusabake, M. & Bishop, J. K. B. ^{226}Ra -derived nutrient budgets in the upper equatorial Pacific and the role of “new” silicate in limiting productivity. *Deep-Sea Res.* **42**, 479–497 (1995).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, New York, 1982).
- Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *J. Phycol.* **21**, 347–357 (1985).
- Dugdale, R. C., Wilkerson, F. P., Barber, R. T. & Chavez, F. P. Estimating new production in the equatorial Pacific Ocean at 150°W . *J. Geophys. Res.* **97**, 681–686 (1992).
- McCarthy, J. J., Garside, C., Nevins, J. L. & Barber, R. T. New production along 140°W in the equatorial Pacific during and following the 1992 El Niño event. *Deep-Sea Res.* **43**, 1065–1093 (1996).
- Probyn, T. A. The inorganic nitrogen nutrition of phytoplankton in the southern Benguela: New production, phytoplankton size, and implications for pelagic food webs. *S. Afr. J. Sci.* **12**, 411–420 (1992).
- Nelson, D. M., Goering, J. J. & Boisseau, D. W. in *Coastal Upwelling* Vol. 1 (ed. Richards, F. A.) (Am. Geophys. Un., Washington DC, 1981).
- Michaels, A. F. & Silver, M. W. Primary production, sinking fluxes and the microbial food web. *Deep-Sea Res.* **35**, 473–490 (1988).
- Wanninkhof, R. A. *et al.* Seasonal and lateral variations in carbon chemistry of surface water in the eastern equatorial Pacific during 1992. *Deep-Sea Res.* **42**, 387–409 (1995).
- Bidigare, R. R. & Ondrusek, M. E. Spatial and temporal variability of phytoplankton pigment distributions in the central equatorial Pacific Ocean. *Deep-Sea Res.* **43**, 809–833 (1996).
- Kaczmarek, I. & Fryxell, G. A. Microphytoplankton of the equatorial Pacific: 140°W meridional transect during the 1992 El Niño. *Deep-Sea Res.* **42**, 535–558 (1995).
- Buesseler, K. O., Andrews, J. A., Hartman, M. C., Belastock, R. & Chai, F. Regional estimate of the export flux of particulate organic carbon derived from Thorium-234 during the JGOFS EqPac program. *Deep-Sea Res.* **42**, 777–804 (1995).
- Barber, R. T. *et al.* Primary productivity and its regulation in the equatorial Pacific during and following the 1991–1992 El Niño. *Deep-Sea Res.* **43**, 933–969 (1996).
- Sunda, W. G., Swift, D. G. & Huntsman, S. A. Low iron requirement for growth in oceanic phytoplankton. *Nature* **351**, 55–57 (1991).
- Fitzwater, S. E., Coale, K. H., Gordon, R. M., Johnson, K. S. & Ondrusek, M. E. Iron deficiency and phytoplankton growth in the equatorial Pacific. *Deep-Sea Res.* **43**, 995–1015 (1996).
- Dugdale, R. C. & Wilkerson, F. P. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 107–122 (Plenum, New York, 1992).
- Archer, D. E. *et al.* Daily, seasonal and interannual variability of sea surface carbon and nutrient concentrations in the equatorial Pacific Ocean. *Deep-Sea Res.* **43**, 779–809 (1996).

Acknowledgements. This work was supported by the US NSF and Region Provence Alpes Maritimes Cooperative Agreement.

Correspondence and requests for materials should be addressed to R.C.D. (e-mail: rdugdale@sfu.edu).

Three-dimensional glacial flow and surface elevation measured with radar interferometry

Johan J. Mohr, Niels Reeh & Søren N. Madsen

Danish Center for Remote Sensing, Department of Electromagnetic Systems, Technical University of Denmark, DK-2800 Lyngby, Denmark

Outlet glaciers—which serve to drain ice from ice sheets—seem to be dynamically less stable in North Greenland than in South Greenland^{1–3}. Storstrømmen, a large outlet glacier in northeastern Greenland which surged between 1978 and 1984 (ref. 2), has been well studied. In general, neither glacier surge mechanisms nor the geographical distribution of the surges are well known. Conventional satellite radar interferometry can provide large-scale topography models with high resolution⁴, and can measure

the radar line-of-sight component of ice-flow vectors⁵, but cannot map full vector flow fields. Here we present an interferometry method that combines observations from descending and ascending satellite orbits which, assuming ice flow parallel to the topographic surface, allows us to use the differing view angles to estimate full three-dimensional surface flow patterns. The accuracy of our technique is confirmed by the good agreement between our radar-based flow model and *in situ* Global Positioning System (GPS) reference data at Storstrømmen. Radar measurements such as these, made regularly and at high spatial density, have the potential to substantially enhance our understanding of glacier dynamics and ice-sheet flow, as well as improve the accuracy of glacier mass-balance estimates.

First visited in 1908⁶, Storstrømmen has a long record of observations. From 1950 to 1978 the glacier front was relatively stable at a retreated position. From 1978 to 1984 the front advanced 12 km, nearly reaching the 1908 position, increasing the glacier area

by 90 km². Fieldwork in the period 1989–95, initiated by the Alfred Wegener Institute⁷, shows that the glacier is now returning to its pre-surge state⁸.

Insight into the dynamics of non-steady (surging) glaciers is essential for proper interpretation of frontal retreat/advance and local ice thinning/thickening. Surges of outlet glaciers may change the topography and mass balance of large ice-sheet regions, thereby contributing significantly to the freshwater input from land ice to the sea. This in turn has an effect on global sea level and on oceanic global circulation.

Satellite interferometric synthetic aperture radar (InSAR) has previously been used for earthquake mapping⁹. Lately, applying InSAR to estimate ice motion and surface elevation has attracted attention. Using external digital elevation models (DEMs), the slant range velocity has been separated from topographic effects^{10,11}. Using multiple interferograms, topography and slant range velocities have been determined simultaneously^{12–15}. Direct comparisons with

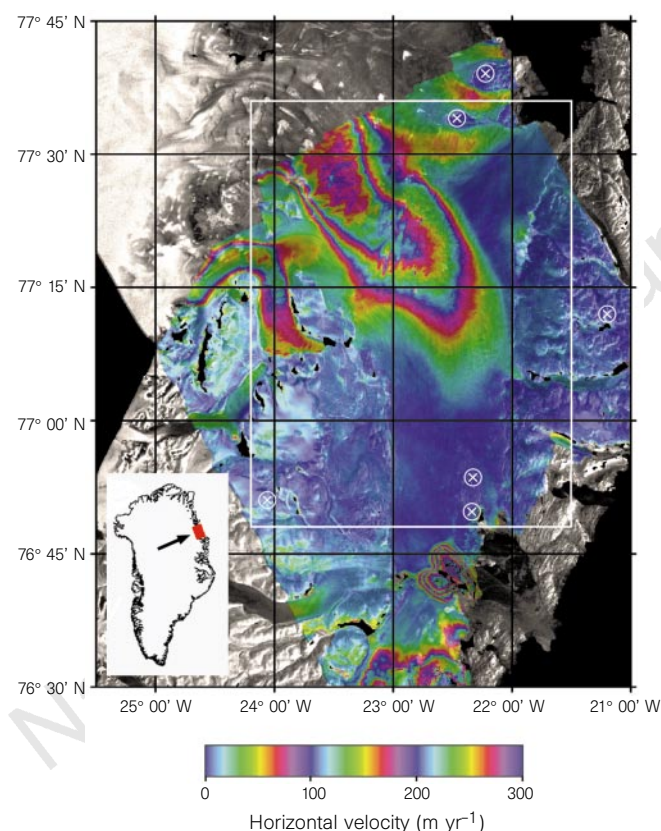


Figure 1 Colour-coded display of annual horizontal displacement in a region around the Storstrømmen glacier. The NNW- and SSW-oriented strips are ascending and descending orbit data respectively. Ice velocities are determined in the 90 km × 90 km overlap region. Descending orbit ERS-1 data were acquired with a 35-d interval on 28 October and 2 December 1995, with ERS-2 repeats 1 d later. The ascending orbit data sets were acquired with a 70-d interval on 31 January and 10 April 1996, with ERS-2 repeats 1 d later. The ambiguity heights (the height differences equivalent to a target moving half a wavelength, one fringe in the interferogram, away from the radar from one acquisition to the next) are approximately –500 m and +5,000 m for the descending, and –70 m and +450 m for the ascending orbit data. After flowing SE at an altitude of 800 m a.s.l. (upper left), the glacier turns south (image centre), with a minor branch continuing to the east towards Kofoed-Hansen Brae (upper right corner). The heart-shaped feature at the front is an artefact caused by tidal motion not being of constant velocity, as our processing assumes. Six “⊗” symbols indicate tie-points used to calibrate heights and velocities. The white box indicates the location of Fig. 2. The northward flow of the Bistrup Brae is just visible at the bottom of the figure.

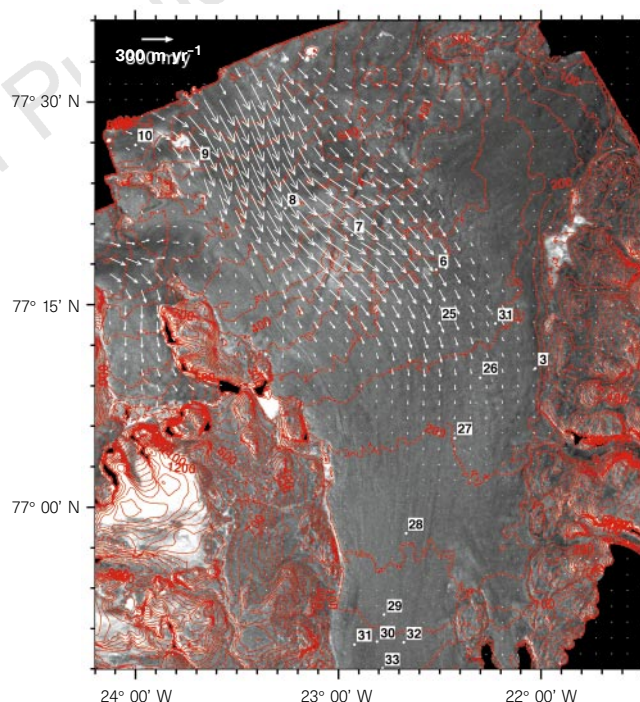


Figure 2 Map (90 km × 70 km) of radar-derived elevations and flow vectors. White dots, and corresponding numbers, show the positions of the stakes listed in Table 1. Surface elevation contours (red) are given at 50-m intervals. The elevations shown are derived from the ascending orbit data. Data have been processed to a pixel spacing of ~100 m × 100 m, but only a small subset of the velocity vectors (white) are displayed. A grey-scale radar amplitude image is shown as background. The slow-down of the flow of Storstrømmen near the terminus is contrary to normal outlet glacier flow, but is a typical post-surge glacier-flow pattern. The smooth, large scale variation of the measured flow field agrees with *in situ* measurements, but often deviates from the local surface gradient. This could not be predicted by glacier dynamics models, demonstrating the value of a direct measurement of both flow magnitude and direction. Using elevation contours for locating the divide separating the ice flowing to Storstrømmen and Kofoed-Hansen Brae (the NE branch of the glacier system), it is expected that the ice flux towards Kofoed-Hansen Brae is ~25% of the total flux of the Storstrømmen/Kofoed-Hansen Brae glacier system. The derived velocity field shows only insignificant ice flux towards Kofoed-Hansen Brae, which was not known to be essentially stagnant.

in situ velocity measurements are scarce^{5,10} and quantitative use of interferometric data has been limited, as the one velocity component derived in previous work is generally insufficient to provide a description of the complicated flow pattern of most outlet glaciers. One velocity component only suffices if the direction of ice flow can be estimated otherwise¹⁰, as in the interior of an ice sheet where ice flow is in the direction of maximum surface slope, or for ice streams confined between mountain walls. In general, knowledge of the full three-dimensional flow pattern is essential for validation of ice-flow models and in particular for understanding the dynamics of non-steady glacier flow.

Here we use interferometric ERS-1/2 tandem mode pairs acquired from both descending and ascending satellite tracks having different 'look' directions to measure the full flow pattern (Fig. 1). For each satellite track (ascending or descending), we used two interferograms to solve for topography and displacement, thus eliminating the need for an independent DEM. The longest ascending baseline (139 m) is more accurate for topographic determination compared with the longest descending baseline (19 m), and thus our final DEM is based on ascending data only. All interferograms are formed from data with one day separation. Data separated by 35 days, or more, decorrelated unacceptably.

Precision satellite state-vector data, with an absolute accuracy of some decimetres (ref. 16), allow us to coregister individual radar images. They also facilitate automatic coregistration with tie-points used to refine the state-vector derived baseline estimates, in effect calibrating the height and velocity measurements derived from the two interferograms. We have used six tie-points on bedrock for the presented results (Fig. 1).

The three-dimensional location and the displacement in the 'look' direction is determined for each pixel. The elevations, displacements, and line-of-sight vectors for each image point for both the descending and ascending data are resampled to geographic coordinates. Adding the assumption that the velocity component in the surface normal direction is zero¹⁴, the three-dimensional velocity vector is estimated (Fig. 2). In a parallel development, Joughin, Kwok and Fahnestock used a similar

North Greenland¹⁷. These workers also conclude that the technique is promising, even though they note that no ground-truth data were available for their study.

A first assessment of the accuracy of the derived velocities is performed by analysing the estimated velocity of bedrock. Small windows in areas with moderate topography show mean deviations up to 5 m yr⁻¹, with a standard deviation of 1.4–2 m yr⁻¹ in the N–S direction and 0.5–0.6 times smaller in the E–W direction. In areas of rough topography, somewhat larger standard deviations are generally observed. The errors consist of high-frequency components typically less than 2 m yr⁻¹, for example, due to thermal noise, and low-frequency errors up to 5 m yr⁻¹, for example, due to tie-point uncertainties, atmospheric distortions, and baseline errors. In one data set an unknown very localized systematic artefact with regular geometry, of magnitude ± 0.5 cycle, is observed in the southern part of a descending pass.

For the Storstrømmen glacier we have the unique possibility of extending the validation with direct measurements of ice flow velocity and elevation. This glacier has been visited for six field seasons during 1989–95, and the positions of 23 stakes drilled into the ice surface were repeatedly measured with GPS (1992–95) and TRANSIT satellite Doppler surveying (1989–90)⁷. The 17 stakes within our study area are shown on a radar map (Fig. 2) with elevation contour lines and a vector overlay of the derived flow. A comparison of the *in situ* data and the radar measurements is summarized in Table 1.

The mean value of the north component of the velocity difference is -2.8 m yr⁻¹ with a standard deviation of 8.7 m yr⁻¹. The corresponding mean and standard deviations of the east component are -3.5 m yr⁻¹ and 4.9 m yr⁻¹. Directions derived from InSAR agree to within a few degrees with the GPS-derived directions for all stakes with velocities greater than ~ 20 m yr⁻¹ (Table 1). This fact strengthens the confidence in the InSAR results, as the direction of glacial flow is expected to show much less variation than the speed of the flow. The differences in the magnitude of the InSAR and the GPS winter velocities are of the same order of magnitude as the variability of the GPS velocities. We conclude that the differences between the InSAR and GPS derived velocities reflect both real

Table 1 Comparison of *in situ* and radar measurements on the Storstrømmen glacier

Stake	Flow magnitude (m yr ⁻¹)					Flow direction (deg)		Elevation (m)	
	Radar	Radar –GPS _w	GPS _w (Winter)	GPS _a (Annual)	GPS _s (Summer)	Radar†	Radar† –GPS	Radar	Radar –GPS
10	20	12	8	8	NA	(154)	(75)	811.5	16
9	248	35	213 ± 5	219 ± 5	248	160	7	674.1	9
8	238	13	225 ± 10	231 ± 10	261	144	3	566.2	2
7	210	–7	217 ± 6	228 ± 6	281	119	4	406.3	10
6	126	–7	133 ± 6	147 ± 6	221	148	4	330.1	14
3.1	38	–15	53 ± 10*	58 ± 10	NA	149	4	299.5	–7
3	13	6	7	7	NA	(149)	(44)	255.1	–8
25	108	8	100 ± 8	113 ± 8	182	159	6	317.9	16
26	41	–4	45 ± 6	63 ± 6	148	175	8	264.0	–17
27	17	7	24 ± 5	37 ± 5	104	204	14	242.8	–4
28	0	–3	3	6	20	(329)	(144)	190.0	5
29	2	1	1	1	NA	(288)	(59)	147.0	–3
30	2	1	1	1	NA	(257)	(83)	115.4	–9
31	1	0	1	1	NA	(293)	(159)	109.6	–5
32	2	1	1	1	NA	(300)	(90)	121.4	–5
33	1	–4	5	5	NA	(208)	(–122)	103.3	–4
34	2	1	1	2	5	(160)	(165)	73.1	14

Table shows comparison of radar measurements of ice flow vectors and elevations with *in situ* data from field seasons 1990–95¹⁸. 'Stake' indicates stakes driven into the glacier (see text). When comparing the magnitude of InSAR- and GPS-derived velocities, the different acquisition times must be considered. The GPS observations are generally performed in the summer at intervals of ~ 1 yr, allowing annual velocities (GPS_a) to be derived. When two or more annual values are available, both the mean annual value and the associated standard deviation are given. The standard deviation encompasses both observation errors and inter-annual variability of the glacier motion. For some locations, velocities covering a two-month period in the summer of 1995 are available. These summer velocities (GPS_s) are between 13% and 180% higher than the mean annual values. To compare GPS and InSAR velocities, GPS winter velocities (GPS_w) are estimated on the assumption that the annual motion is composed of two months (the approximate length of the melt season) of fast summer motion and 10 months of relatively slow, but constant, winter motion. As the ERS data used are acquired in the winter season, there is no contradiction between this assumed annual velocity variation and the basic assumption of constant flow velocity during the InSAR observation period. Due to the track angles of 332° and 208° for ascending and descending passes respectively, the noise component of the radar data is expected to be smaller in the E–W direction than in the N–S direction, by a factor of $\tan 28^\circ \approx 0.53$, in agreement with observed deviations. Stake 9, located in an area with large velocity and elevation gradients, is not included in our statistics. NA, not available.

* Winter GPS-velocity is estimated as GPS-annual value reduced by 10%.

† In these columns, radar directions and deviations of points with GPS-velocities less than 20 m yr⁻¹ are in parentheses.

moderate, inter-annual variations and measurement errors. This conclusion is supported by the much smaller r.m.s. deviation of only 2.5 m yr^{-1} for the InSAR/GPS velocity differences of the points located in the stagnant part of the glacier, where the contribution to the uncertainty from flow variability is negligible.

The absolute accuracy of the elevations is of the order of 10 m, compared to the GPS measurements. Because the GPS points lie in fairly flat areas and the local r.m.s. noise in the radar DEM is of the order of a metre, the observed elevation errors are primarily caused by low-frequency errors.

Observed accuracies are consistent with errors expected from atmospheric artefacts, the constant velocity, and the surface parallel flow assumptions (discussed below). A 0.5-cm atmospheric delay perturbation (one-way) will, with baselines and time separations of this study, lead to a 10-m elevation error if applied to one of the ascending interferograms and a 5 m yr^{-1} horizontal velocity error if applied to the ascending or descending interferograms with the smallest topographic sensitivity. In polar regions, typical dynamic atmospheric one-way delay perturbations are expected to be $<0.5 \text{ cm}$. Perturbations of similar magnitude can be caused by dry uneven snow layers. When a pair of baselines used to determine both elevation and velocity differ significantly, the velocity estimated is approximately that of the shorter spatial baseline. Here the descending velocity is that of 2–3 December 1995, the ascending velocity is that of 10–11 April 1996; both are winter velocities, expected to vary little over the period. For comparison, thermal noise is of the order of 1–2 mm differential range on each interferogram, equivalent to height accuracies of the order of 1–2 m. Observed correlations were generally high on the glacier, with average values of 0.8–0.9. An uncertainty in the slope of 1° translates into a relative velocity uncertainty of $\sim 4\%$.

The discrepancies between the InSAR and GPS measurements are of the order of the variability observed in the GPS data. However, there are several possibilities for improving the InSAR technique. The technique used here relies on an assumption of surface-parallel flow which even for stationary glaciers is not completely valid, because the ice particles at the surface move downwards in the accumulation zone, and upwards in the ablation zone. As the Storstrømmen glacier is not stationary, an additional vertical flow component is present. For Storstrømmen, the total vertical motion not accounted for by the assumption of surface-parallel flow should be less than $3\text{--}4 \text{ m yr}^{-1}$, causing a maximum error on the horizontal velocity of $\sim 10 \text{ m yr}^{-1}$. However, if the ice thickness was known, the surface-parallel-flow approximation could be replaced by the correct condition derived by using the principle of mass conservation (the continuity equation). Using redundant data to minimize the influence of atmospheric perturbations would further improve the accuracy, and three-dimensional flow velocities with an absolute accuracy of $1\text{--}2 \text{ m yr}^{-1}$ seem within reach.

The InSAR measurements give new important information about the surge and post-surge phases of Storstrømmen. Comparing the GPS measurements of surface elevations with elevations obtained from aerial photographs from 1978 (that is, before the surge) shows that ice has been transferred from the central to the lower part of the glacier⁸. The InSAR area-covering measurements enable us to calculate the ice volumes involved in this mass transfer. At present, the lower part of the glacier is almost stagnant and is therefore thinning due to summer melting. The InSAR maps also reveal that Kofoed-Hansen Brae, the northeastern branch of Storstrømmen (Fig. 2) for which no other measurements are available, is at present a stagnant ice body. Two stagnant ice plugs thus prevent ice from the central part of Storstrømmen from being removed, and summer melting cannot balance the upward movement. This part of the glacier is therefore building up. Today, InSAR height measurements are not accurate enough to directly detect this process, but combining detailed InSAR velocity measurements with ice thickness and ablation data, by means of the continuity equation, will allow

calculation of surface elevation (ice volume) changes. The availability of one more 'look' direction would allow direct measurement of vertical velocities. The application of simultaneous measurements of heights and three-dimensional velocities is not restricted to surging glaciers like Storstrømmen. It is a new and powerful method for assessing volume changes of land-ice masses. $\square$

Received 6 February; accepted 20 October 1997.

1. Mock, S. J. Fluctuation of the terminus of the Harald Moltke Brae, Greenland. *J. Glaciol.* **6**, 369–373 (1966).
2. Reeh, N., Bøggild, C. E. & Oerter, H. Surge of Storstrømmen, a large outlet glacier from the Inland Ice of North-East Greenland. *Rapp. Grønlands. Geol. Unders.* **162**, 201–209 (1994).
3. Joughin, I., Tulaczyk, S., Fahnestock, M. & Kwok, R. A mini-surge on the Ryder Glacier, Greenland, observed by satellite radar interferometry. *Science* **274**, 228–230 (1996).
4. Goldstein, R. M., Zebker, H. A. & Werner, C. L. Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Sci.* **23**, 713–720 (1988).
5. Goldstein, R. M., Engelhardt, H., Kamb, B. & Frolich, R. M. Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* **262**, 1525–1530 (1993).
6. Koch, J. P. & Wegener, A. Die Glaziologischen Beobachtungen der Danmark-Expedition. *Meddr. Grønland* **46**, 1–77 (1911).
7. Oerter, H., Bøggild, C. E., Jung-Rothenhäuser, F. & Reeh, N. in *Climate and Sea Level Change and the Implications for Europe* Appendix A, 2–6 (Final Rep., Contract EV5V-CT91-0051, Commission of the European Communities DGXII, Brussels, 1995).
8. Reeh, N., Bøggild, C. E., Jung-Rothenhäuser, F. & Oerter, H. in *Climate and Sea Level Change and the Implications for Europe* Appendix D, 2–17 (Final Rep., Contract EV5V-CT91-0051, Commission of the European Communities, DGXII, Brussels, 1995).
9. Massonnet, D. *et al.* The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* **364**, 138–142 (1993).
10. Rignot, E., Jezek, K. C. & Sohn, H. G. Ice flow dynamics of the Greenland ice-sheet from SAR interferometry. *Geophys. Res. Lett.* **22**, 575–578 (1995).
11. Joughin, I. R., Winebrenner, D. P. & Fahnestock, M. A. Observations of ice-sheet motion in Greenland using satellite radar interferometry. *Geophys. Res. Lett.* **22**, 571–574 (1995).
12. Kwok, R. & Fahnestock, M. A. Ice sheet motion and topography from radar interferometry. *IEEE Trans. Geosci. Remote Sensing* **34**, 189–200 (1996).
13. Joughin, I., Winebrenner, D., Fahnestock, M., Kwok, R. & Krabill, W. Measurement of ice-sheet topography using satellite-radar interferometry. *J. Glaciol.* **42**, 10–22 (1996).
14. Joughin, I., Kwok, R. & Fahnestock, M. Estimation of ice-sheet motion using satellite radar interferometry: Method and error analysis with application to Humboldt Glacier, Greenland. *J. Glaciol.* **42**, 564–575 (1996).
15. Rignot, E. Tidal motion, ice velocity and melt rate of Petermann Gletscher, Greenland, measured from radar interferometry. *J. Glaciol.* **42**, 476–485 (1996).
16. Massman, F. H., Neymayer, J. C., Raimondo, K., Enninghorst & Li, H. Quality of the D-PAF orbits before and after the inclusion of PRARE data. in *Proc. 3rd ERS Symp.* Vol. 3, 1655–1660 (SP-414, ESA, Noordwijk, 1998).
17. Joughin, I., Kwok, R. & Fahnestock, M. Interferometric estimation of three-dimensional ice-flow using ascending and descending Passes. *IEEE Trans. Geosci. Remote Sensing* (in the press).
18. Bøggild, C. E., Jung-Rothenhäuser, F. & Oerter, H. Glaciological investigations on Storstrømmen glacier, North-East Greenland. *Express Report, Grønlands. Geol. Unders.* 163–170 (Geol. Survey of Denmark and Greenland, Copenhagen, 1995).

Acknowledgements. The Danish Center for Remote Sensing is funded by the Danish National Research Foundation. ERS data have been made available by the European Space Agency. Field work on Storstrømmen was supported by the Commission of the European Communities.

Correspondence and requests for materials should be addressed to S.N.M. (e-mail: snm@emi.dtu.dk).

Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression

Gerald S. Wilkinson, Daven C. Presgraves* & Lili Crymes

Department of Zoology, University of Maryland, College Park, Maryland 20742, USA

* Department of Biology, University of Rochester, Rochester, New York 14627, USA

In some species, females choose mates possessing ornaments that predict offspring survival^{1–5}. However, sexual selection by female preference for male genetic quality^{6–8} remains controversial because conventional genetic mechanisms maintain insufficient variation in male quality to account for costly preference and ornament evolution^{9,10}. Here we show that females prefer ornaments that indicate genetic quality generated by transmission conflict between the sex chromosomes. By comparing sex-ratio distributions in stalk-eyed fly (*Cyrtodiopsis*) progeny we found that female-biased sex ratios occur in species exhibiting eye-stalk sexual dimorphism^{11,12} and female preferences for long eye span^{13,14}. Female-biased sex ratios result from meiotic drive¹⁵,

the preferential transmission of a 'selfish' X-chromosome. Artificial selection for 22 generations on male eye-stalk length in sexually dimorphic *C. dalmanni* produced longer eye-stalks and male-biased progeny sex ratios in replicate lines. Because male-biased progeny sex ratios occur when a drive-resistant Y chromosome pairs with a driving X chromosome¹⁵, long eye span is genetically linked to meiotic drive suppression. Male eye span therefore signals genetic quality by influencing the reproductive value of offspring¹⁶.

In Malaysia, two closely related species, *C. dalmanni* and *C. whitei*, exhibit biased sex ratios with 33–35% males¹⁷ and form nocturnal aggregations in which male mating success increases with male eye span^{13,14}. Although females possess eye-stalks, the slope of the regression of eye-stalk on body length in males is greater than in females, causing extreme sexual dimorphism for eye span in both species (Fig. 1a, b). Field observations at night indicate that females move further (1.16 ± 0.23 m) between aggregation sites on successive nights than males (0.47 ± 0.20 m; analysis of variance (ANOVA): $F_{1,50} = 5.16$, $P = 0.028$), and males often return to the same site on successive nights. Female-choice experiments using pairs of males differing only in eye span, resulting from artificial selection in *C. dalmanni*¹⁴ or physical manipulation in *C. whitei*¹³, confirm that females prefer to roost and mate¹⁸ with males with long eye-stalks. In contrast, *C. quinqueguttata* did not exhibit sexual dimorphism in eye span (Fig. 1c) or sex-ratio bias (53% males, $n = 197$). *C. quinqueguttata* do not aggregate at night¹¹, and in mate-choice experiments females show no preference for males with larger eye span or body size (G.S.W. and H. Kahler, unpublished data).

To determine whether variation in sex ratio between these congeneric species is consistent with a driving X chromosome, X^d , we compared sex ratios in progeny obtained from individual males.

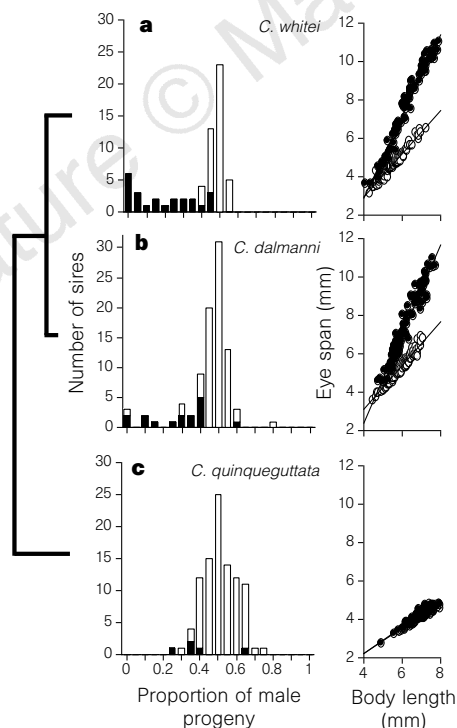


Figure 1 Sex-ratio distribution of male stalk-eyed flies, individuals were of the species: **a**, *Cyrtodiopsis whitei*; **b**, *C. dalmanni*; **c**, *C. quinqueguttata*. Shaded bars indicate males with progeny sex ratios that deviated (χ^2 test, adjusted for continuity; $P < 0.05$) from 0.5. Phylogenetic relationships are indicated on the left. Right, eye span on body length least-squares regressions for field-captured flies of each species. Filled circles, males; open circles, females.

The frequency of biased sex ratio was 36% in *C. whitei* ($n = 75$ sires, 10,395 progeny; Fig. 1a) after about five generations in culture and 16% in *C. dalmanni* ($n = 93$ sires, 10,695 progeny; Fig. 1a) after approximately 30 generations in culture. These X^d frequencies do not differ from estimates¹⁵ obtained for field-captured flies (contingency tests: *C. dalmanni*, $\chi^2 = 0.14$, $P = 0.71$; *C. whitei*, $\chi^2 = 0.05$, $P = 0.82$). In contrast, *C. quinqueguttata* males showed no consistent progeny sex-ratio bias ($n = 102$ sires, 3,553 progeny; Fig. 1c). Phylogenetic reconstruction based on over 3 kilobases of mitochondrial and nuclear sequences indicates that *C. quinqueguttata* is basal to the other two *Cyrtodiopsis* species and that eye-stalk monomorphism is plesiomorphic in the family Diopsidae (R. Baker and G.S.W., manuscript in preparation). Thus, exaggerated male eye-stalks, female preferences and a stable X-linked meiotic drive polymorphism apparently evolved after divergence from *C. quinqueguttata*.

To establish that there is a genetic relationship between exaggerated male eye span and chromosome drive we measured eye span and sex ratios of *C. dalmanni* flies selected for ratio of eye span to body length in males¹⁹. For each generation we selected 10 of 50 males to mate with 25 unselected females in replicate high, low and control lines. After 22 generations, artificial sexual selection primarily changed eye span (Fig. 2). By ANOVA, 95% of eye-span variation could be explained using sex ($F_{1,438} = 6966.2$, $P < 0.0001$), selection treatment ($F_{2,438} = 385.8$, $P < 0.0001$) and replicate ($F_{1,438} = 7.8$, $P < 0.006$) as factors, but they account for only 37% of body-length variation. For body length, the effect of replicate ($F_{1,438} = 30.4$, $P < 0.001$) was greater than selection ($F_{2,438} = 16.1$, $P < 0.001$).

The distribution of biased progeny sex ratios also differed between selection lines (Kruskal–Wallis test: $H = 47.2$, $P < 0.0001$). Compared to the *C. dalmanni* stock population (Fig. 1a), one line with decreased eye span exhibited no change in median sex ratio ($n = 107$ sires, 7,801 flies, Mann–Whitney U test: $Z = 0.44$, $p = 0.66$; Fig. 3a), whereas males from the other low line produced more female-biased progeny ($n = 110$ sires, 8,327 flies, $Z = 2.91$ sires, $P = 0.0036$; Fig. 3b). In contrast, males from both lines selected for increased eye span produced fewer female-biased and more male-biased progeny than stock population males ($n = 111$ sires, 10,652 flies, $Z = 2.90$, $P = 0.0016$, Fig. 3c; $n = 101$ sires, 8,721 flies, $Z = 3.16$, $P = 0.0013$, Fig. 3d). We have

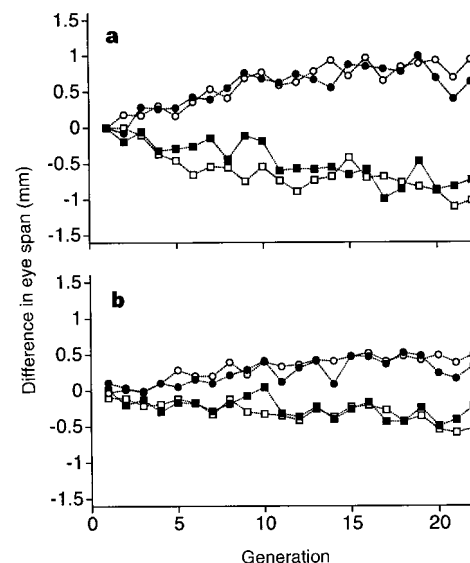


Figure 2 Replicate bidirectional response to selection on male eye-span to body-length ratio plotted against generation for eye span expressed as differences from control line averages. **a**, Males; **b**, females.

previously shown¹⁵ that a modifying Y chromosome, Y^m , causes 63% male-biased progeny sex ratios when paired with X^d . The form of spermiogenesis disruption in sex-ratio biasing *C. dalmanni* males suggests that male excess occurs because X^d damages some X^d -bearing, but not Y^m -bearing sperm, in X^dY^m males¹⁵. Therefore, we estimate X^d frequency within each line as the fraction of sires producing either female-biased (X^dY) or male-biased (X^dY^m) progenies, and Y^m frequency as the proportion of sex-ratio-biasing males producing male-biased progenies, that is $X^dY^m/(X^dY + X^dY^m)$. Given a stock population frequency estimate of 6.7% and 10 males mating in each of 22 generations, the probability that Y^m should increase in frequency by drift, only in the high lines, is 0.005 (Monte Carlo simulation). Mean male eye span from each line correlates significantly with the frequency of Y^m (Fig. 4b) but not X^d (Fig. 4a), indicating associated Y-linked effects on drive suppression and eye span.

Meiotic drive has been implicated in the evolution of female preferences in at least one other system. Female house mice heterozygous for transmission-distorting t-haplotypes prefer $+/+$ over $+/t$ males and avoid producing inviable t/t offspring²⁰. The benefits of choosing long eye span in stalk-eyed flies differ because sex-chromosome drive distorts the sex ratio. Consequently, choosy females produce more sons in a female-biased population, and those sons carry genes for both long eye-stalks and meiotic drive suppression. Choosy females also enhance the fertility of male offspring because 48% of X^dY , but only 11% of X^dY^m , male sperm degenerate¹⁵. The selective forces maintaining the X^d polymorphism have not yet been identified. However, the low frequency of Y^m

suggests that suppression is costly¹⁵, as expected for a stable polymorphism^{21,22}. Recent studies report drive suppressors on the Y chromosome and every autosome in *Drosophila simulans*²³, as well as polymorphism for Y-linked drive suppression in *D. mediopunctata*²² and *D. paramelanica*²⁴. If sex-chromosome meiotic drive is indeed more widely distributed than suspected^{25,26}, then other cases of sex-linked indicators of drive suppression should occur. For example, some guppy populations exhibit sex-ratio meiotic drive²⁷ and female preferences for males with Y-linked colour patterns^{28,29}. To the extent that X-linked drivers, or other costly selfish genetic elements, continuously coevolve with suppressors³⁰, genetic variation for fitness should persist and allow male ornament exaggeration. □

Methods

Field work. We captured flies near streams 20–32 km north of Kuala Lumpur, Malaysia, in January 1989 and 1996. To estimate daily movements we netted 47 *C. dalmanni* and 55 *C. whitei* along a mountain stream and marked individuals with a unique combination of paint spots on the thorax. Then, in 14 consecutive nights, we recorded location to the nearest 0.05 m for all marked flies roosting along 25 m of stream. We calculated the average distance each fly moved per day between the first and last day sighted.

Sex ratio. For each species and selected line we mated six or more females to each male for a week and then allowed isolated females to oviposit on 50 ml of medium¹⁹ for 2 weeks. We estimated sex ratios for each male by pooling the male and female progeny produced by all of his mates. To obtain sufficient virgin flies for breeding from each selected line we used flies after 22 generations of selection or a succeeding generation in which 25 males and 25 females from each line were arbitrarily selected for mating. We detected no change in mean eye span within each line between these generations.

Received 14 July; accepted 22 October 1997.

- Norris, K. Heritable variation in a plumage indicator of viability in male great tits, *Parus major*. *Nature* **362**, 537–539 (1993).
- Møller, A. P. Male ornament size as a reliable cue to enhanced offspring viability in the bar swallow. *Proc. Natl Acad. Sci. USA* **91**, 6929–6932 (1994).
- Petrie, M. Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature* **371**, 598–599 (1994).
- Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229–232 (1996).
- Sheldon, B. C., Merilä, J., Qvarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
- Pomiankowski, A. The evolution of female mate preferences for male genetic quality. *Oxf. Surv. Evol. Biol.* **5**, 136–184 (1988).
- Heywood, J. S. Sexual selection by the handicap mechanism. *Evolution* **43**, 1387–1397 (1989).
- Iwasa, Y., Pomiankowski, A. & Nee, S. The evolution of costly mate preferences. II. The “handicap” principle. *Evolution* **45**, 1431–1442 (1991).
- Burt, A. Perspective: the evolution of fitness. *Evolution* **49**, 1–8 (1995).
- Kirkpatrick, M. & Barton, N. H. The strength of indirect selection on female mating preferences. *Proc. Natl Acad. Sci. USA* **94**, 1282–1286 (1997).
- Burkhardt, D. & de la Motte, I. Selective pressures, variability, and sexual dimorphism in stalk-eyed flies (Diptera: Diopsidae). *Naturwissenschaften* **72**, 204–206 (1985).
- Wilkinson, G. S. & Dodson, G. Function and evolution of antlers and eyestalks in flies. in *The Evolution of Mating Systems in Insects and Arachnids* (eds Choe, J. & Crespi, B.) (Cambridge Univ. Press, 1997).
- Burkhardt, D. & de la Motte, I. Big ‘antlers’ are favoured: female choice in stalk-eyed flies (Diptera, Insecta), field collected harem and laboratory experiments. *J. Comp. Physiol. A* **162**, 649–652 (1988).
- Wilkinson, G. S. & Reillo, P. R. Female preference response to artificial selection on an exaggerated male trait in a stalk-eyed fly. *Proc. R. Soc. Lond. B* **255**, 1–6 (1994).
- Presgraves, D. C., Severance, E. & Wilkinson, G. S. Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* **147**, 1169–1180 (1997).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, New York, 1958).
- Burkhardt, D. & de la Motte, I. How stalk-eyed flies eye stalk-eyed flies: observations and measurements of the eyes of *Cyrtodiopsis whitei* (Diptera). *J. Comp. Physiol.* **151**, 407–421 (1983).
- Burkhardt, D., de la Motte, I. & Lunau, K. Signalling fitness: larger males sire more offspring. Studies of the stalk-eyed fly *Cyrtodiopsis whitei* (Diptera). *J. Comp. Physiol. A* **174**, 61–64 (1994).
- Wilkinson, G. S. Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera: Diopsidae). *Genet. Res.* **62**, 213–222 (1993).
- Lenington, S. The t complex: a story of genes, behavior, and populations. *Adv. Study Behav.* **20**, 51–86 (1991).
- Wu, C.-I. The fate of autosomal modifiers of the sex-ratio trait in *Drosophila* and other Sex linked meiotic drive systems. *Theor. Popul. Biol.* **24**, 107–120 (1983).
- Carvalho, A. B., Vaz, S. C. & Klaczko, L. B. Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of *Drosophila mediopunctata*. *Genetics* **146**, 891–902 (1997).
- Montchamp-Moreau, C., Cazemajor, M. & Landré, C. The sex ratio trait in *Drosophila simulans*: genetic analysis of distortion and suppression. *Genetics* (in the press).
- Stalker, H. D. The genetic systems modifying meiotic drive in *Drosophila paramelanica*. *Genetics* **46**, 177–202 (1961).
- Hurst, L. & Pomiankowski, A. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane’s rule and related phenomenon. *Genetics* **128**, 841–858 (1991).

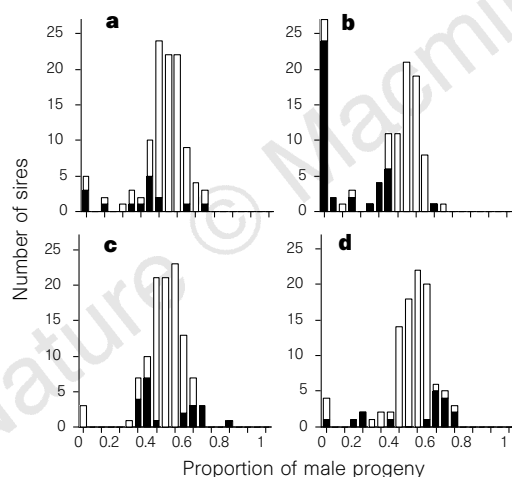


Figure 3 Sex-ratio distributions of male *C. dalmanni* after 22 generations of selection. Selection was replicated for decreased (a, b) or increased (c, d) male relative eye span. Shaded bars denote biased sex ratios, as in Fig. 1.

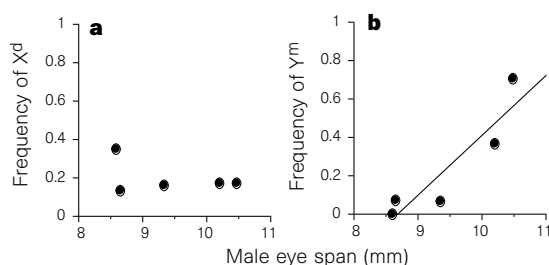


Figure 4 Correlation between mean male eye span and chromosome frequency. a, Driving X chromosome, X^d ($r = 0.39$, $P = 0.49$); b, drive-resistant Y chromosome, Y^m ($r = 0.91$, $P = 0.02$). Frequency among the stock population and four selected lines after 22 generations of selection is shown.

26. Jaenike, J. Sex-ratio meiotic drive in the *Drosophila quinaria* group. *Am. Nat.* **148**, 237–254 (1996).
27. Farr, J. A. Biased sex ratios in laboratory strains of guppies, *Poecilia reticulata*. *Heredity* **47**, 237–248 (1981).
28. Houde, A. E. Mate choice based upon naturally occurring color-pattern variation in a guppy population. *Evolution* **41**, 1–10 (1987).
29. Houde, A. E. Sex-linked heritability of a sexually-selected character in a natural population of guppies, *Poecilia reticulata* (Pisces: Poeciliidae). *Heredity* **69**, 229–235 (1992).
30. Hurst, L. D. Further evidence consistent with Stellate's involvement in meiotic drive. *Genetics* **142**, 641–643 (1996).

Acknowledgements. We thank H. S. Yong, P. Reillo and M. Taper for assistance in the field; S. Bhatt, P. Reillo and D. Reames for assistance in the laboratory; and C. Boake, R. Gibson, L. Hurst, M. Kirkpatrick, R. Trivers and L. Wolfenbarger for comments on the manuscript. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to G.S.W. (e-mail: wilkinson@zool.umd.edu).

Spontaneous sign systems created by deaf children in two cultures

Susan Goldin-Meadow & Carolyn Mylander

University of Chicago, Department of Psychology, 5730 South Woodlawn Avenue, Chicago, Illinois 60637, USA

Deaf children whose access to usable conventional linguistic input, signed or spoken, is severely limited nevertheless use gesture to communicate^{1–3}. These gestures resemble natural language in that they are structured at the level both of sentence⁴ and of word⁵. Although the inclination to use gesture may be traceable to the fact that the deaf children's hearing parents, like all speakers, gesture as they talk⁶, the children themselves are responsible for introducing language-like structure into their gestures⁷. We have explored the robustness of this phenomenon by observing deaf children of hearing parents in two cultures, an American and a Chinese culture, that differ in their child-rearing practices^{8–12} and in the way gesture is used in relation to speech¹³. The spontaneous sign systems developed in these cultures shared a number of structural similarities: patterned production and deletion of semantic elements in the surface structure of a sentence; patterned ordering of those elements within the sentence; and concatenation of propositions within a sentence. These striking similarities offer critical empirical input towards resolving the ongoing debate about the 'innateness' of language in human infants^{14–16}.

We videotaped four deaf children in the USA (one in Philadelphia and three in Chicago) and four in Taiwan, Republic of China (Taipei), interacting with their hearing mothers at home with a standard set of toys. Each child was observed twice between 3 years 8 months and 4 years 11 months. Children were congenitally deaf with no recognized cognitive deficits. Cause of the deafness was unknown. Each child had at least a 70 to 90 dB hearing loss in both ears; even with hearing aids, none was able to acquire speech naturally. Children attended oral schools advocating training in sound sensitivity, lip-reading and speech production. At the time of videotaping, none could do more than produce an occasional spoken word in a highly constrained context.

None of the children had been exposed to a conventional sign system. The children's hearing parents attempted to communicate with them through speech. However, much of their interaction took place in action and gesture, a technique that worked because conversation was about the 'here-and-now'.

Although the number of subjects in the study was small, the number of observations was not. The data contain 6,614 gestural communications (an average of 455 per child, 372 per mother) made up of 10,398 individual gestures (779 per child, 531 per mother).

Unlike hearing children and adults who rarely concatenate their

spontaneous gestures into strings^{17,18}, the deaf children in both cultures often conveyed their message through gesture sentences rather than single gestures. The sentences the Chinese and American children produced conformed closely to a structural analogue of the ergative pattern that predominates in some, but not all, natural languages^{19,20}; importantly, not in English or Mandarin. This observation reduces the likelihood that the grammatical structure noted in the deaf children's gesture systems was somehow derived from the structure of the spoken languages that surrounded them.

The hallmark of an ergative pattern is that the actor in an intransitive sentence (mouse in the proposition 'mouse goes to hole') is distinguished linguistically from the actor in a transitive sentence (mouse in 'mouse eats cheese') and instead is marked like the patient (cheese). In contrast, in English, which is predominantly an accusative language, intransitive actors are treated like transitive actors and not like patients; for example, both actors precede the verb ('the mouse goes to the hole' and 'the mouse eats the cheese') whereas patients follow the verb ('the mouse eats the cheese').

Children in both cultures produced gestures for transitive actors, patients and intransitive actors at different rates (Fig. 1, dark bars; $F(2, 7) = 22.52$, $P < 0.0001$ for the groups combined; proportional data subjected to Freeman–Tukey transform before statistical analysis²¹). Gestures were produced significantly more often for patients (eaten-cheese) and for intransitive actors (moving-mouse) than for transitive actors (eating-mouse; both comparisons $P < 0.01$, Newman–Keuls). There were no significant differences between patients and intransitive actors. This production prob-

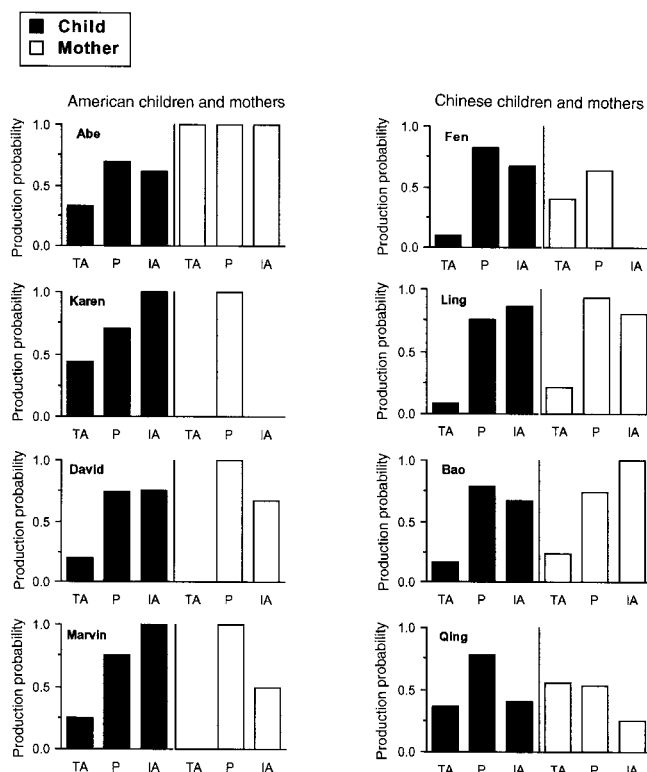


Figure 1 Probability that a gesture will be produced for transitive actors (TA), patients (P), and intransitive actors (IA) in a two-element gesture sentence. Probabilities were calculated using sentences in which three semantic elements could be gestured but only two elements actually were gestured. Deaf children (dark bars) showed significant differences in production patterns across the three elements. Gestures were produced more often for patients than for transitive actors, and more often for intransitive actors than for transitive actors, a structural analogue of the ergative pattern found in certain natural languages. Hearing mothers (white bars) were not consistent in their treatment of intransitive actors and did not display an ergative pattern.

ability conformed to an ergative pattern: gesture production was high and equal for intransitive actors and patients, and low for transitive actors. Note that production probability patterns convey information about who is the doer and the done-to in two-element sentences. If, for example, a deaf child produced the gesture sentence 'boy hit', it is likely that the boy was the hittee (patient) rather than the hitter (transitive actor) in the scene under description.

In addition to producing reliably some semantic elements at the expense of others, children were also consistent in where those elements were positioned in two-gesture strings. Children in both cultures produced gestures for intransitive actors before gestures for acts (*mouse-go*); Ling produced 14 of 15 relevant sentences conforming to this pattern, Qing 17 of 19, Bao 12 of 15, David 18 of 23, Abe 10 of 10 (P values ≤ 0.02 ; binomial test on each child), Marvin 4 of 4, Karen 2 of 3 (Fen was an exception, 1 of 4). Children also produced gestures for patients before gestures for acts (*cheese-eat*); Ling 11 of 12, Bao 26 of 29, Fen 9 of 11, Qing 29 of 29, Marvin 9 of 10 (P values ≤ 0.03), Abe 9 of 14, Karen 4 of 4 (David was an exception, 17 of 35; the precocious onset of a noun/verb distinction may have perturbed this pattern which was evident in his earlier sessions²²). Thus, children in both cultures placed intransitive actors in the same position as patients, a pattern again consistent with an ergative structure.

Only two children, Qing and David, produced gestures for transitive actors often enough for order to be explored. David produced gestures for transitive actors after gestures for acts (*eat-mouse*, 5 of 6; this order has been confirmed in larger samples²³). David thus distinguished transitive from intransitive actors, not only in production probability, but also in gesture order. Qing produced gestures for transitive actors in the same position as intransitive actors and patients: before gestures for acts (*mouse-eat*, 8 of 8, $P \leq 0.004$). However, Qing did distinguish between patients and transitive actors when the two appeared in the same string, producing gestures for transitive actors in second position after patients (*cheese-mouse*, 6 of 7, $P \leq 0.06$). Patients and intransitive actors thus consistently occupied first position in two-element sentences, again conforming to an ergative structure.

Children in both cultures produced complex sentences, concatenating more than one proposition within a single string. For example, one child produced a 'clap' gesture, a point at himself, a 'twist' gesture, a 'blow' gesture, and a point at mother to request mother to twist open the jar (proposition₁) and blow a bubble (proposition₂) so that he could clap it (proposition₃). The children did not differ significantly in their production of complex sentences

($U = 5$, NS, Mann-Whitney U Test²⁴; Fen 0.54 of 65 gesture sentences; Bao 0.42 of 289; Qing 0.41 of 257; Ling 0.33 of 156; David 0.45 of 394; Abe 0.36 of 152; Karen 0.34 of 50; Marvin 0.32 of 82). Each child's gesture system thus demonstrated generative capability, an essential property of all natural languages.

We next explored a potential source of input to the children's gesture systems, analysing the spontaneous gestures produced by the hearing mothers when interacting with their children during these videotaped sessions. Like their children, mothers produced transitive actors, patients, and intransitive actors at significantly different rates (Fig. 1, white bars; $F(2, 6) = 4.20$, $P < 0.05$). Karen's mother produced no relevant intransitive sentences and was consequently excluded from statistical analysis ($n = 7$); all other blank entries in the figure reflect the fact that the mothers did not produce a semantic element when they had the opportunity to do so (on average, they had 17.3 such opportunities). Mothers produced more gestures for patients than for transitive actors ($P < 0.05$), thus distinguishing between the two, as did their children. It is, however, where intransitive actors are situated relative to transitive actors and patients that determines the typology of a language, and here mothers and children differed. Mothers showed no reliable patterning of intransitive actors, whereas children produced intransitive actors at a rate significantly different from transitive actors but not different from patients, thus displaying an ergative pattern.

Chinese mothers tended to order their gestures within sentences in the same way as their children (patient-act, mothers Bao 16 of 21, Fen 14 of 16, Ling 11 of 11, P values < 0.01 ; intransitive actor-act, mothers Bao 5 of 5, Fen 7 of 7, Qing 6 of 6, P values < 0.02 , Ling 9 of 12, $P < 0.07$), with the exception that Qing's mother showed no patient-act order (6 of 12) and no patient-transitive actor order (2 of 5) although her child did. In contrast, American mothers each produced only 2 to 5 relevant gesture sentences in these sessions, and more extensive analyses of additional sessions²³ also failed to reveal statistically significant ordering patterns.

Across the two cultures, mothers produced proportionally fewer complex sentences than their children ($t(7) = 2.50$, $P < 0.05$ on transformed data²¹; Fig. 2; total $n = 1,445$ children, 629 mothers), although the gap was much wider for American dyads than for Chinese. To explore who took the lead in the production of complex gesture sentences, we looked at the developmental pattern for American mothers and children in additional videotaped sessions (Fig. 3; total $n = 1,364$ children, 229 mothers). In 15 of 20 sessions

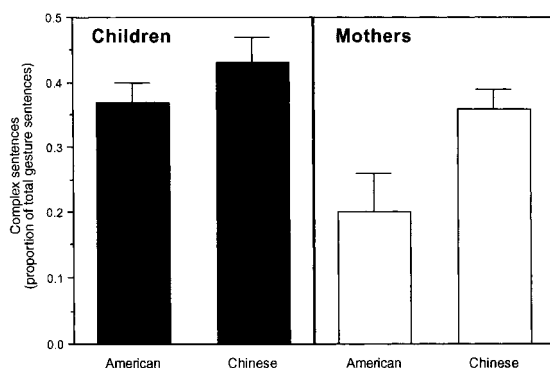


Figure 2 Mean number of complex gesture sentences as a proportion of total gesture sentences in Chinese and American deaf children and their mothers. Complex sentences contain two or more propositions. The children in the two cultures did not differ in their production of complex sentences and, as a group, produced significantly more complex sentences than their mothers. The error bars represent s.e.m.

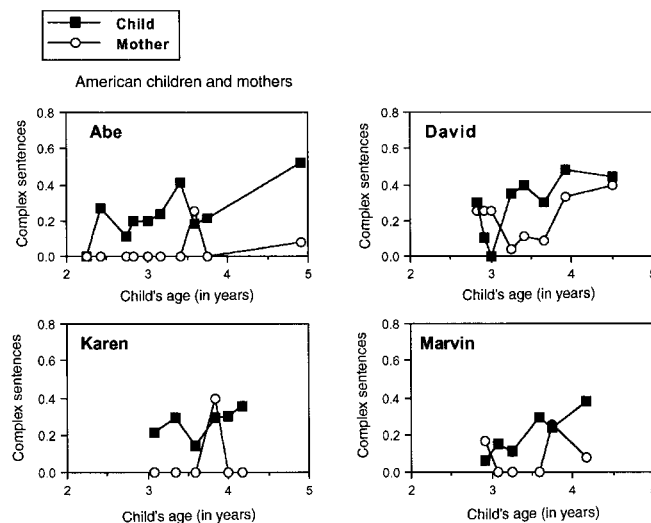


Figure 3 Development of complex gesture sentences in American deaf children and their mothers (proportion of total gesture sentences). All four children steadily increased their use of complex sentences over time. The mothers did not consistently use complex sentences until months after their children, if at all.

before age 3 years 8 months, child produced proportionally more complex gesture sentences than mother ($P < 0.02$, sign test²⁴). In only 6 of the 20 sessions did mothers' production of complex sentences rise above 4%, compared to 18 of 20 for the children ($z = 3.55$, $P < 0.001$). Thus, American children were unlikely to have learned complex gesture sentences from their mothers.

Overall, American mothers' gestures did not resemble their children's. Indeed, American children's gestures had more in common with Chinese children's gestures than with their own mothers'. American children thus appear to be responsible for the structural aspects of their systems. In contrast, Chinese mothers' gestures did resemble their children's, at least in part. Chinese children may therefore have learned segments of their systems from their mothers or, more likely given that Chinese and American children's gestures follow the same patterns, the mothers may have learned them from their children. If so, we ask why Chinese (but not American) mothers copy gesture patterns from their deaf children. The answer might involve cultural differences in attitudes toward children's communications, or the languages themselves (it may be easier to produce complex gesture sentences while speaking Mandarin than while speaking English). Whatever the reason, the fact remains that American children took the lead in creating their gesture systems, a lead their mothers did not follow.

Given the salient differences between Chinese and American cultures^{8–12}, the structural similarities in the children's gesture systems are striking. These structural properties—consistent marking of semantic elements by deletion and by ordering, and concatenation of propositions within a single sentence—are developmentally robust in humans (but not, apparently, in chimpanzees²⁵). Their development is buffered against large variations in environmental conditions and in this sense can be considered 'innate'²⁶. □

Methods

Children's and mothers' gestures were coded according to a system developed previously²³. Criteria for isolating gestures grew out of a concern that the gestures meet the minimal requirements for a communicative symbol: first, the gesture must be directed to another individual; the gesturer must establish eye contact with a communication partner, or be assured of the partner's attention, before acting; and second, gesture must not itself be a direct manipulation of some relevant person or object; it must be empty-handed¹⁷. Using these criteria, we isolated gestures from the stream of motor behaviour. We characterized the form of the gestures following guidelines established for coding conventional sign languages, and divided gestures into sentence strings on the basis of motoric criteria. We then characterized the meaning of the gestures, deciding how many and what type of propositions were conveyed in a sentence, and identifying individual semantic elements.

Reliability was established between two trained coders who independently transcribed a portion of the videotapes. Two native Mandarin speakers (one born and raised in Taiwan) who were bilingual in English coded the Chinese videotapes, and two native English speakers coded the American videotapes. Reliability was 87% and 89% agreement between coders (for the Chinese and American samples, respectively) for describing gesture form, 100% and 93% for identifying sentence boundaries, 87% and 85% for identifying types of propositions, and 92% and 90% for identifying semantic elements.

Received 25 April; accepted 20 October 1997.

1. Moores, D. F. in *Language Perspectives* (eds Schiefelbusch, R. L. & Lloyd, L. L.) 377–378 (University Park, Baltimore, 1974).
2. Tervoort, B. T. Esoteric symbolism in the communication behavior of young deaf children. *Am. Ann. Deaf* **106**, 436–480 (1961).
3. Fant, L. *Ameslan* (National Association of the Deaf, Silver Springs, MD, 1972).
4. Goldin-Meadow, S. & Feldman, H. The development of language-like communication without a language model. *Science* **197**, 401–403 (1977).
5. Goldin-Meadow, S., Mylander, C. & Butcher, C. The resilience of combinatorial structure at the word level: morphology in self-styled gesture systems. *Cognition* **56**, 195–262 (1995).
6. Feyereisen, P. & deLannoy, J.-D. *Gesture and speech* (Cambridge Univ. Press, New York, 1991).
7. Goldin-Meadow, S. & Mylander, C. Gestural communication in deaf children: the non-effects of parental input on language development. *Science* **221**, 372–374 (1983).
8. Miller, P. J., Wiley, A., Fung, H. & Liang, C.-H. Personal storytelling as a medium of socialization in Chinese and American families. *Child Development* **68**, 557–568 (1997).
9. Stevenson, H. W. et al. Contexts of achievement. *Monogr. Soc. Res. Child Dev.* **55** 1–119 (1990).

10. Young, N. F. Socialization patterns among the Chinese of Hawaii. *Amerasia J.* **1**, 31–51 (1972).
11. Chen, C. & Uttal, D. H. Cultural values, parents' beliefs, and children's achievement in the United States and China. *Hum. Dev.* **31**, 351–358 (1988).
12. Wang, X.-L., Mylander, C. & Goldin-Meadow, S. in *Research on Family Resources and Needs Across the World* (ed. Cusinato, M.) 363–374 (Edizioni Universitarie di Lettere Economia, Diritto, Italy, 1996).
13. McNeill, D. & Duncan, S. D. in *Language and Gesture: Window into Thought and Action* (ed. McNeill, D.) (New Lexington, San Francisco, submitted).
14. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA, 1965).
15. Pinker, S. *The Language Instinct* (Harper Collins, New York, 1995).
16. Gleitman, L. R. & Newport, E. L. in *Language, Vol. 1, Invitation to Cognitive Science Series* (eds Gleitman, L. R. & Liberman, M.) 1–24 (MIT Press, Cambridge, MA, 1995).
17. Petitto, L. A. in *The Development of Language and Language Researchers* (ed. Kessel, F.) 187–221 (Erlbaum Associates, Hillsdale, 1988).
18. McNeill, D. *Hand and Mind* (Univ. Chicago Press, Chicago, 1992).
19. Dixon, R. M. W. *Ergativity* (Cambridge Univ. Press, New York, 1994).
20. Silverstein, M. in *Grammatical Categories in Australian Languages* (ed. Dixon, R. M. W.) 112–171 (Australian Institute of Aboriginal Studies, Canberra, 1976).
21. Zar, J. H. *Biostatistical Analysis* 2nd edn (Prentice-Hall, Englewood Cliffs, 1984).
22. Goldin-Meadow, S., Butcher, C., Mylander, C. & Dodge, M. Nouns and verbs in a self-styled gesture system: what's in a name? *Cogn. Psychol.* **27**, 259–319 (1994).
23. Goldin-Meadow, S. & Mylander, C. Gestural communication in deaf children: the effects and non-effects of parental input on early language development. *Monogr. Soc. Res. Child Dev.* **49**, 1–151 (1984).
24. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, 1956).
25. Goldin-Meadow, S. Book review of Kanzi: *The Ape at the Brink of the Human Mind* by Savage-Rumbaugh, S. & Lewin, R. *Int. J. Primatol.* **17**, 145–148 (1996).
26. Wimsatt, W. in *Integrating Scientific Disciplines* (ed. Bechtel, W.) 185–208 (Martinus-Nijhoff, Dordrecht, 1986).

Acknowledgements. We thank L. Barsalou, J. Huttenlocher, B. Keysar, M. K. McClintock, P. Miller, D. McNeill and W. Wimsatt for comments; S. Duncan for help in collecting the Taiwanese data; and F. Bechter and M.-Y. Zheng for help in transcribing the videotapes. Supported by grants from the National Institute on Deafness and other Communication Disorders and the NSF.

Correspondence and requests for materials to S.G.-M. (e-mail: sgscg@ccp.uchicago.edu).

Prostaglandins stimulate calcium-dependent glutamate release in astrocytes

Paola Bezzi*, Giorgio Carmignoto†, Lucia Pasti†, Sabino Vesce*, Daniela Rossi*, Barbara Lodi Rizzini*, Tullio Pozzan† & Andrea Volterra*

* Institute of Pharmacological Sciences, University of Milan, Via Balzaretti 9, 20133 Milan, Italy

† Department of Experimental Biomedical Sciences and CNR Center for the Study of Biomembranes, University of Padova, Viale G. Colombo 3, 35121 Padova, Italy

Astrocytes in the brain form an intimately associated network with neurons. They respond to neuronal activity and synaptically released glutamate by raising intracellular calcium concentration ($[Ca^{2+}]_i$)^{1,2}, which could represent the start of back-signalling to neurons^{3–5}. Here we show that coactivation of the AMPA/kainate and metabotropic glutamate receptors (mGluRs) on astrocytes stimulates these cells to release glutamate through a Ca^{2+} -dependent process mediated by prostaglandins. Pharmacological inhibition of prostaglandin synthesis prevents glutamate release, whereas application of prostaglandins (in particular PGE_2) mimics and occludes the releasing action of GluR agonists. PGE_2 promotes Ca^{2+} -dependent glutamate release from cultured astrocytes and also from acute brain slices under conditions that suppress neuronal exocytotic release. When applied to the CA1 hippocampal region, PGE_2 induces increases in $[Ca^{2+}]_i$ both in astrocytes and in neurons. The $[Ca^{2+}]_i$ increase in neurons is mediated by glutamate released from astrocytes, because it is abolished by GluR antagonists. Our results reveal a new pathway of regulated transmitter release from astrocytes and outline the existence of an integrated glutamatergic cross-talk between neurons and astrocytes *in situ* that may play critical roles in synaptic plasticity and in neurotoxicity.

The release of endogenous glutamate from cultured cortical astrocytes was monitored continuously by means of an enzymatic assay⁶. Coapplication of (S)- α -amino-3-hydroxy-5-methyl-4-

isoxazolepropionic acid (AMPA) and (1S,3R)-1-aminocyclopentane-1,3-dicarboxylic acid (*t*-ACPD), agonists of AMPA/kainate and mGluRs, respectively, induced potent and rapid release of glutamate from the astrocytes (Fig. 1a) in a dose-dependent manner (Fig. 1b). *t*-ACPD alone, but not AMPA alone, also stimulated the release, although to a much lower extent. Glutamate release was mediated by AMPA receptors (AMPARs) and mGluRs linked to the inositol-1,4,5-trisphosphate (Ins(1,4,5)P₃)–Ca²⁺ cascade⁷ because: (1) it was potentiated by cyclothiazide and reduced by GYKI 52466, two agents selective for the AMPAR subtypes; (2) stimulation with the selective group I mGluR agonist (*R,S*)-3,5-dihydroxyphenylglycine (DHPG) was as effective as *t*-ACPD, whereas groups II (L-CCGI) and III (L-AP4) mGluR agonists were much less effective (Fig. 1c).

Astrocytes can release glutamate through multiple mechanisms, both Ca²⁺-independent (reversed glutamate transport⁸ and swelling-evoked release⁹) and Ca²⁺-dependent⁴. The effect of AMPA + *t*-ACPD was: (1) insensitive to the inhibitory action of dihydro-

kainate (2–5 mM) on astrocyte glutamate transport¹⁰; (2) additive to glutamate release through the transporters induced by *t*-pyrrolidine-2,4-dicarboxylic acid (PDC)¹¹ (+83.5 ± 8.4%; *n* = 3; Fig. 1d); and (3) not associated with astrocyte swelling (98.8 ± 0.5% of the fluorescence emitted from unstimulated cells by the dye 2'7'-bis(2-carboxyethyl)-5(6)-carboxy fluorescein/AM at its isosbestic point¹² (*n* = 3). In contrast, it was drastically reduced either by stimulating in a medium without Ca²⁺ or upon increasing the intracellular buffering capacity by cell loading with the Ca²⁺-chelator BAPTA/AM. Substitution of external Na⁺ with *N*-methyl-D-glucamine (NMDG) also decreased the release (Fig. 1c). When bradykinin was used to evoke Ca²⁺-dependent release, coadministration of AMPA + *t*-ACPD did not produce further stimulation (+6 ± 2%, *n* = 3; Fig. 1d). The release process induced by AMPA + *t*-ACPD in analogy with that of bradykinin^{4,13}, was sensitive to the anion transport inhibitor furosemide (–71 ± 3.5%; *n* = 7) and to tetanus neurotoxin (TeNT, 2 µg ml^{–1}), a blocker of neuronal exocytotic release¹⁴ (Fig. 1c). Importantly, TeNT inhibition developed slowly: first observed after 8 h of exposure (18.5 ± 1.2%, *n* = 3), it progressively increased to 72.5 ± 5% (*n* = 6) after 20 h. In parallel, release through the glutamate transporters with PDC was unaffected (not shown). We conclude that the release mechanism activated by GluR agonists in astrocytes is totally distinct from swelling- or transporter-mediated release and it apparently shares common properties with the Ca²⁺-dependent release process elicited by bradykinin^{13,15}.

In striatal neurons, associative activation of AMPARs and mGluRs stimulates phospholipase A₂ (PLA₂) and arachidonic acid release¹⁶. We investigated whether the same mechanism plays a role in the glutamate release process of astrocytes. Indeed: (1) AMPA + *t*-ACPD (and *t*-ACPD alone, to a lower extent) rapidly enhanced ³H-arachidonic acid liberation through a Ca²⁺-dependent

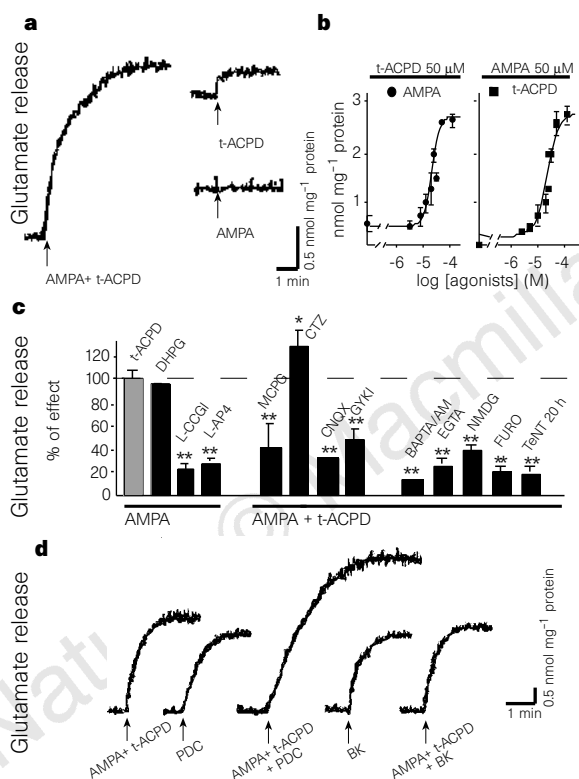


Figure 1 Ca²⁺-dependent glutamate release from astrocytes in response to joint stimulation of AMPARs and mGluRs. **a**, Fluorescence traces representing glutamate release elicited by AMPA and *t*-ACPD together (each at 50 µM) or individually (at 100 µM). **b**, Dose-response curves of AMPA + *t*-ACPD-evoked release: each agonist curve is in the presence of 50 µM of the co-agonist. Points represent mean ± s.e.m. (*n* ≥ 4); **c**, release induced by 50 µM AMPA + 50 µM *t*-ACPD (=100%, striped bar) and effect of different pharmacological manipulations. Left, Substitution of *t*-ACPD with subtype-selective mGluR agonists: (*R,S*)-3,5-dihydroxyphenylglycine (DHPG, 100 µM); (2S,1'S,2'S)-2-(carboxycyclopropyl)-glycine (L-CCGI, 1 µM); *L*(+)-2-amino-4-phosphonobutyric acid (L-AP4, 150 µM). Middle, Addition of cyclothiazide (CTZ, 100 µM) or GluR antagonists: (*R,S*)-α-methyl-4-carboxyphenylglycine (MCPG, 500 µM) for mGluRs; 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 10 µM) and GYKI 52466 (GYKI, 50 µM) for AMPA/kainate receptors. Right, Pretreatment of cultures with 1,2-bis(2-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid tetrakis(acetoxymethyl)ester (BAPTA/AM, 50 µM), furosemide (FURO, 5 mM) or tetanus neurotoxin for 20 h (TeNT 20h); stimulation in a medium free of Ca²⁺ (EGTA) or Na⁺ (NMDG). **P* < 0.05; ***P* < 0.01, one-way ANOVA and Scheffé method for multiple comparisons (*n* = 3–8). **d**, Additivity of the glutamate-releasing action of AMPA + *t*-ACPD (each at 100 µM) to that of PDC (2 mM) but not to that of bradykinin (BK) (100 nM).

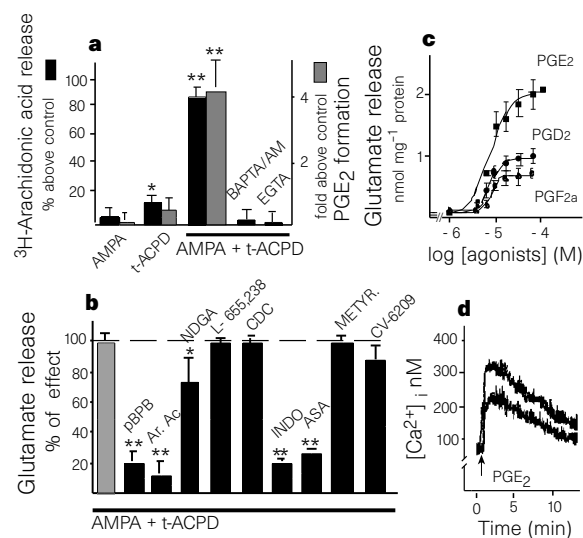


Figure 2 Activation of the arachidonate cascade and involvement of prostaglandins in the Ca²⁺-dependent release of glutamate from astrocytes. **a**, Liberation of ³H-arachidonic acid-derived material (left scale, solid bars) and formation of PGE₂ (right scale, striped bars) in response to AMPA, *t*-ACPD or AMPA + *t*-ACPD (all at 50 µM); preloading of the cells with BAPTA/AM (50 µM) or stimulation in Ca²⁺-free medium (EGTA) prevents the response: *n* = 5–9; **P* < 0.05; ***P* < 0.01, one-way ANOVA and Scheffé method for multiple comparisons. **b**, Glutamate release evoked by AMPA + *t*-ACPD in the absence (striped bar = 100%) and presence of arachidonic acid cascade inhibitors (solid bars; *n* = 3–10; statistics as in **a**). **c**, Dose-response curves of glutamate release stimulated by PGE₂, D₂ and F_{2a}. **d**, [Ca²⁺]_i elevations induced by PGE₂ (20 or 50 µM) in populations of astrocytes¹⁹. At 20 µM, PGE₂ increased [Ca²⁺]_i (92 ± 4 nM, *n* = 82) by 145 ± 15 nM (*n* = 6); at 50 µM, by 251 ± 31 nM (*n* = 7); its effects in **c** and **d** were resistant to 1 µM indomethacin (not shown).

action (peak at 1 min; Fig. 2a and legend for details); and (2) inhibitors of the eicosanoid-forming enzymes¹⁷ affected AMPA + *t*-ACPD-evoked glutamate release (Fig. 2b). In particular, release was drastically reduced by blocking either PLA₂ (with *p*-bromophenacylbromide (pBPB, 10 μ M) or aristolochic acid (Ar.Ac., 50 μ M)) or cyclooxygenase (with indomethacin (INDO, 1 μ M) or aspirin (ASA, 10 μ M)). In contrast, inhibitors of lipoxygenases (nordihydroguaiaretic acid (NDGA, 10 μ M), L-655,238 (1 μ M) or cinnamyl-3,4-dihydroxy- α -cyanocinnamate (CDC,

0.3 μ M)), of epoxigenase (metyrapone, 100 μ M) and of platelet aggregating factor (PAF) receptors (CV-6209, 1 μ M) were practically ineffective. Prostaglandins D₂, E₂ and F_{2 α} , the major cyclooxygenase products of astrocytes¹⁸, all rapidly stimulated glutamate release (Fig. 2c). PGE₂, which is formed in response to AMPA + *t*-ACPD (Fig. 2a), was the most potent, with a plateau effect approaching that of the GluR agonists. Moreover, when coapplied, the effect of the two stimuli was not additive ($+8 \pm 2\%$ of the effect of PGE₂ alone; $n = 3$), suggesting that they activate a common pathway. Interestingly, the releasing action of PGE₂, like that of AMPA + *t*-ACPD, was prevented by cell loading with BAPTA/AM ($-80 \pm 9\%$; $n = 6$). In addition, PGE₂ (1–50 μ M) produced marked elevation of [Ca²⁺]_i in astrocytes loaded with the dye Fura-2/AM¹⁹ (Fig. 2d). Most of this response ($82.6 \pm 5.3\%$; $n = 7$) was insensitive to the blocking of GluRs by a cocktail of mGluR, AMPAR and NMDAR antagonists (1 mM (*R,S*)- α -methyl-4-carboxyphenylglycine (MCPG), 50 μ M 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and 50 μ M 2-amino-5-phosphonopentanoic acid (D-AP5)). Therefore, PGE₂ mediates glutamate release from astrocytes in response to GluR stimulation by triggering Ca²⁺-dependent events that are essential for release.

We next investigated whether the same phenomenon is expressed by astrocytes from acute cortical or hippocampal slices preincubated for 40 min in TeNT²⁰ (100 μ g ml⁻¹). This treatment abolished neuronal exocytosis but, at least in culture, left Ca²⁺-dependent release from astrocytes nearly intact (-9.2 ± 2.7 ; $n = 4$). The suppression of synaptic glutamate release by TeNT was confirmed by: (1) inhibition of excitatory post-synaptic currents from CA1 neurons evoked by Schaffer collateral stimulation (not shown); (2) abolishment of the massive glutamate release upon high K⁺ stimulation otherwise observed in control slices ($-98 \pm 2\%$; $n = 7$; Fig. 3a, see also b and c). Both AMPA + *t*-ACPD and PGE₂ stimulated glutamate release from TeNT-treated slices (Fig. 3b and c, upper panels) with comparable efficacy in the cortex and hippocampus (not shown). As in cultured astrocytes, indomethacin inhibited the effect of AMPA + *t*-ACPD ($-68.8 \pm 8\%$; $n = 4$; Fig. 3b) and BAPTA that of PGE₂ ($-61.2 \pm 6\%$; $n = 4$; Fig. 3c).

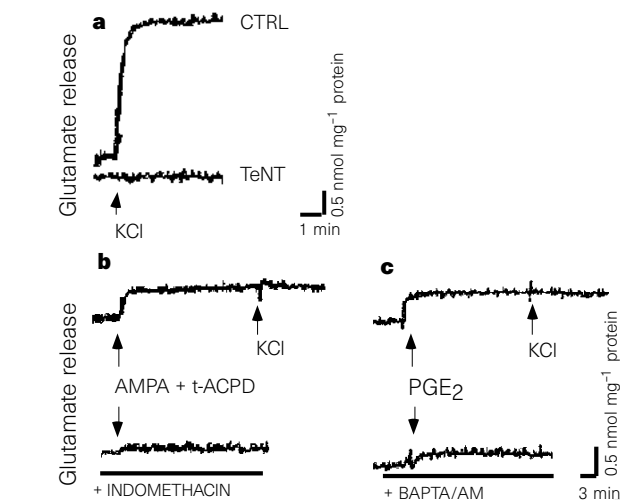


Figure 3 AMPA + *t*-ACPD and PGE₂ stimulate glutamate release from hippocampal slices with blocked neuronal exocytosis. **a**, Glutamate release evoked by high KCl (20–40 mM) in naive slices (CTRL; 2.2 ± 0.2 nmol mg⁻¹ protein, $n = 4$) and in TeNT-treated slices (100 μ g ml⁻¹, 40 min, $n = 7$). **b** and **c**, all slices pre-incubated in TeNT: lack of response to KCl confirms effective blockade of synaptic glutamate release; upper and lower traces, each representative of 4–7 determinations, are from separate slices. **b**, Release evoked by AMPA + *t*-ACPD (each at 50 μ M, upper trace) and inhibition by indomethacin (3 μ M, lower trace). **c**, Response to PGE₂ (20 μ M, 0.48 ± 0.1 nmol mg⁻¹ protein, $n = 6$, upper trace), attenuated in slices loaded with BAPTA/AM (50 μ M, lower trace).

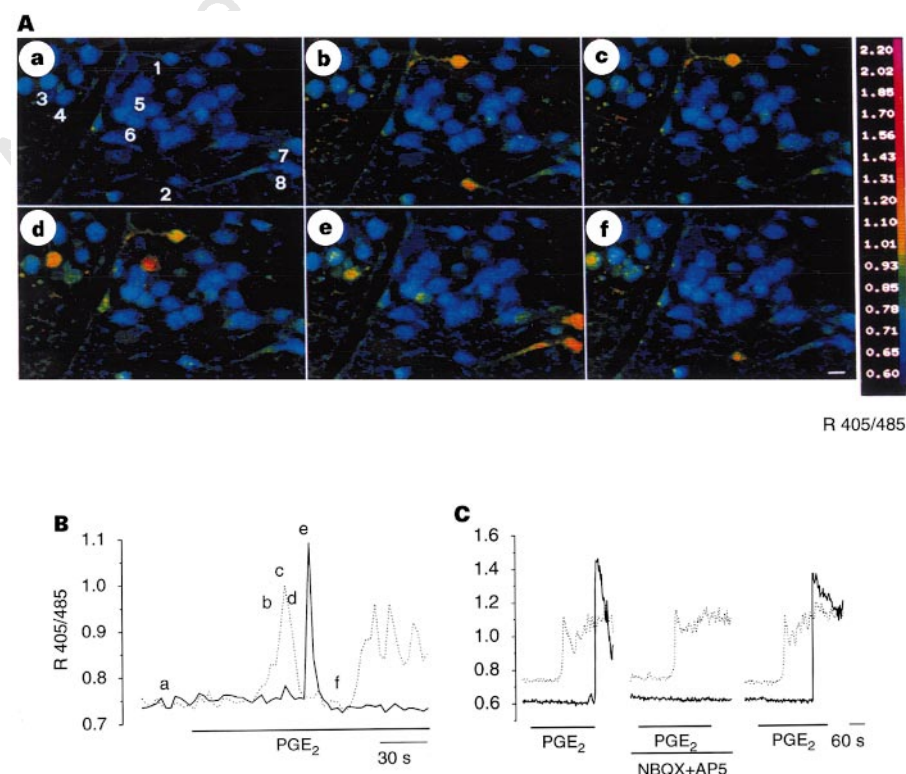


Figure 4 Effects of PGE₂ on the [Ca²⁺]_i of Indo-1-loaded hippocampal cells. **A**, Pseudocolour images of the [Ca²⁺]_i changes in the hippocampal CA1 region of a 7-day-old rat. The sequence shows [Ca²⁺]_i transients in astrocytes 1 and 2 and pyramidal neurons 3–8 upon slice perfusion with 10 μ M PGE₂. Four additional neurons in the field showed delayed [Ca²⁺]_i elevations (not shown) in correspondence to the second [Ca²⁺]_i peak in the astrocytes (see trace in **B**). Acquisition rate, 2 s. Scale bar, 10 μ m. **B**, Kinetics of PGE₂-induced [Ca²⁺]_i changes in the astrocyte 1 (dotted line) and in the neuron 8 (continuous line). Letters a–f correspond to the value of R405/485 as measured from cells in the images a–f in **A**. **C**, [Ca²⁺]_i increases in one pyramidal neuron (continuous line) and one adjacent astrocyte (dotted line) following three consecutive PGE₂ (10 μ M) applications. In the presence of NBQX and D-AP5 (both at 50 μ M) the PGE₂-induced [Ca²⁺]_i increase in the neuron was abolished, whereas that of the astrocyte was unchanged. Acquisition rate: 2 s.

To address the possible physiological significance of glutamate release by astrocytes, we then studied the effects of PGE₂ on the [Ca²⁺]_i of astrocytes and neurons from the CA1 hippocampal region of acute brain slices after incubation with the fluorescent Ca²⁺ indicator Indo-1/AM and TeNT. In the confocal images, CA1 neurons loaded with Indo-1 could be clearly distinguished by their typical pyramidal shape and large soma size, whereas astrocytes were distinguished by their small size, stellate shape and, occasionally, end feet processes that typically contact blood vessels (Fig. 4A, see cell labelled 1). In patch-clamp experiments, more than 90% of these cells revealed features consistent with those of astrocytes, that is, absence of action potential discharges in response to depolarizing current pulses, highly negative resting potentials and dye coupling upon Indo-1 filling through the patch pipette.

PGE₂ induced significant [Ca²⁺]_i increases in astrocytes, such as in cells labelled 1 and 2 in Fig. 4A, and in a number of neurons, such as cells labelled 3–8 (Fig. 4B). In a series of similar experiments, 67% of the pyramidal neurons and 76% of the astrocytes loaded with Indo-1 responded to PGE₂ (0.5–50 μM). To investigate whether the [Ca²⁺]_i increases in neurons and astrocytes were due to glutamate release and activation of GluRs, PGE₂ was applied in the presence of 6-nitro-7-sulphamoyl benzo[*f*]quinoxaline-2,3-dione (NBQX) and D-AP5, antagonists of ionotropic GluRs. Although in most astrocytes (27/33) the response to PGE₂ was unchanged, in a large number of neurons (31 out of 58, *n* = 8) it was abolished by NBQX and D-AP5, but recovered after washout (Fig. 4C). These data demonstrate that the [Ca²⁺]_i increase induced by PGE₂ in neurons is mediated by glutamate, most probably released from astrocytes. This conclusion is strengthened by the observation that, in the presence of either MCPG (1 mM), NBQX and D-AP5 or MCPG alone, the response to PGE₂ was blocked in all the responsive neurons (*n* = 22 and 14, respectively). In the presence of NBQX, D-AP5 and MCPG, the astrocyte response was also affected: 7 out of the 19 PGE₂-responding cells failed to respond, whereas in the remaining astrocytes PGE₂ still largely increased [Ca²⁺]_i (86.3 ± 9.1% of the response amplitude without antagonists; *n* = 12). Likewise, with MCPG alone, 6 out of 12 astrocytes failed to respond to PGE₂ whereas in the remaining astrocytes the amplitude of the [Ca²⁺]_i increase was only slightly reduced (86.8 ± 4.5% of the response without MCPG; *n* = 6). These results suggest that the PGE₂-induced [Ca²⁺]_i elevation in astrocytes depends on two sequential processes: the first one is directly activated by PGE₂, and the second one follows the release of glutamate and the stimulation of mGluRs on the astrocytes themselves. Apparently, this positive feedback mechanism, while present also in culture, is of major relevance *in situ*.

Our results show that astrocytes respond to joint stimulation of the AMPARs and mGluRs by releasing glutamate through a new Ca²⁺-dependent process mediated by prostaglandins. Indeed, the potent and direct elevation of astrocyte [Ca²⁺]_i induced by PGE₂ is apparently essential for activation of release. The mechanism by which PGE₂ increases [Ca²⁺]_i and the steps that specifically link [Ca²⁺]_i rise to transmitter release await further elucidation. We also demonstrate that glutamate released from astrocytes upon PGE₂ stimulation triggers [Ca²⁺]_i elevations in a large number of neurons from the CA1 hippocampal region. As [Ca²⁺]_i rises are crucial in a variety of neuronal functions, including phenomena such as long-term potentiation or depression, the astrocyte modulation of the [Ca²⁺]_i in neurons might disclose an unexpected participation of glial cells in the plasticity of synaptic transmission. Moreover, because GluRs on astrocytes are activated by synaptically released glutamate^{1,2,21,22}, the present data outline the existence of continuous bidirectional communication between neurons and astrocytes, based on a reciprocal glutamatergic signalling²². This communication may integrate with classic synaptic transmission in the processing of information in the brain. Finally, as prostaglandins can be products of altered central nervous system functions^{23–25} and are

formed by glia in response to pro-inflammatory agents²⁶, the prostaglandin-mediated glutamate release from astrocytes may also play a pathophysiological role in a number of brain diseases or injuries. □

Methods

Tissue preparations. Confluent monolayers of cultured astrocytes from the cerebral cortex of newborn rats and transverse thin slices (150–250 μm) either from the visual cortex or the hippocampus of 7–12-day-old rats were prepared as described^{11,27}. At least 97% of the cultured cells were positive to an antibody against the astrocyte-specific glial fibrillary acidic protein (GFAP). Cultures depleted of microglia retained responses to AMPA + *t*-ACPD and PGE₂.

Enzymatic assay of endogenous glutamate release. Efflux of endogenous glutamate was monitored in continuous culture using an enzymatic assay⁶ with modifications. Cultured astrocytes plated on glass coverslips or acute slices gently fixed to a mesh holder were lodged in a 1 × 1 cm cuvette (2 ml volume) inside a Perkin-Elmer LS50B computerized spectrofluorometer at 37 °C under stirring in (in mM): NaCl 120, KCl 3.1, NaH₂PO₄ 1.25, HEPES–Na 25, glucose 4, MgCl₂ 1, CaCl₂ 2 at pH 7.4, added with glutamate dehydrogenase (GDH, 40 U ml^{−1}; Sigma G2626, batch 64H7130) and 1 mM NADP⁺. Glutamate released from the preparations was immediately oxidized by GDH to α-ketoglutarate (thereby preventing the re-uptake process¹⁰) with formation of NADPH and fluorescence emission at 430 nm (delay <1 s; excitation light 335 nm). Release was quantified referring to standard curves constructed with exogenous glutamate and by normalizing for the protein content of each sample. Agents were added directly in the cuvette through a microsyringe, except those acting intracellularly, that required pre-incubation (30 min, unless otherwise specified).

Monitoring of ³H-arachidonic acid release and PGE₂ formation. Release of ³H-arachidonic acid-derived material from cultures pre-incubated overnight with ³H-arachidonic acid (1 μCi ml^{−1}; 209 Ci mmol^{−1}; Amersham) and exposed to stimulants for different times (often 1 min) was monitored according to ref. 28. Data were expressed as percentage increase above basal release (in 1 min = 0.85 ± 0.18% of the incorporated radioactivity; *n* = 7). PGE₂ formation was detected using a highly sensitive and specific ¹²⁵I-radioimmunoassay²⁶ (Amersham kit RPA 530). Cell supernatants were collected after 3-min incubations with stimulants and the data expressed as fold increase above basal level (= 4.5 ± 0.6 pg mg^{−1} protein, *n* = 5).

Confocal fluorescence microscopy. Slices were incubated in the acetoxymethyl derivative of Indo-1 (Indo-1/AM, 20 μM; Molecular Probes, Eugene, OR, USA), 0.02% pluronic acid and purified TeNT (100 μg ml^{−1}) added to the physiological saline (in mM): NaCl 120, KCl 3.1, NaH₂PO₄ 1.25, NaHCO₃ 25, dextrose 4, MgCl₂ 2, CaCl₂ 1, Na-pyruvate 2, Myo-inositol 0.75, ascorbic acid 0.1, at pH 7.4 with 5% CO₂, 95% O₂. Incubation was performed at 37 °C for 40 min under continuous mild stirring. Slices were then mounted in a chamber and placed on the stage of a Nikon inverted microscope (Diaphot 300), equipped with a ×40 immersion objective (NA = 1.1) connected with a real-time confocal system (RCM8000). The ratio of the intensity of the light emitted at the two wavelengths (405/485) was displayed as a pseudocolour scale. During recordings at room temperature, slices were superfused (3 ml min^{−1}) with (in mM): NaCl 120, KCl 3.1, NaH₂PO₄ 1.25, NaHCO₃ 25, dextrose 5, MgCl₂ 1, CaCl₂ 2, at pH 7.4 with 5% CO₂, 95% O₂.

Patch-clamp recordings. Standard procedures for pipette preparation and whole-cell patch-clamp recordings in slices were used^{27,29}. Indo-1 was included in the patch pipette at 500 μM. Schaffer collaterals were stimulated with 50 μs pulses (50 to 200 μA at 0.2 Hz) delivered by a bipolar tungsten electrode (5 μm tip, Roboz, Maryland).

Received 16 September; accepted 20 October 1997.

1. Dani, J. W., Chernjavski, A. & Smith, S. J. Neuronal activity triggers calcium waves in hippocampal astrocyte networks. *Neuron* **8**, 429–440 (1992).
2. Porter, J. T. & McCarthy, K. D. Hippocampal astrocytes *in situ* respond to glutamate released from synaptic terminals. *J. Neurosci.* **16**, 5073–5081 (1996).
3. Nedergaard, M. Direct signaling from astrocytes to neurons in cultures of mammalian brain cells. *Science* **263**, 1768–1771 (1994).
4. Parpura, V. *et al.* Glutamate-mediated astrocyte-neuron signalling. *Nature* **369**, 744–747 (1994).
5. Hassinger, T. D. *et al.* Evidence for glutamate-mediated activation of hippocampal neurons by glial calcium waves. *J. Neurobiol.* **28**, 159–170 (1995).
6. Nicholls, D. G., Sihra, T. S. & Sanchez-Prieto, J. Calcium-dependent and -independent release of glutamate from synaptosomes monitored by continuous fluorimetry. *J. Neurochem.* **49**, 50–57 (1987).

7. Steinhäuser, C. & Gallo, V. News on glutamate receptors in glial cells. *Trends Neurosci.* **19**, 339–345 (1996).
8. Szatkowski, M., Barbour, B. & Attwell, D. Non-vesicular release of glutamate from glial cells by reversed electrogenic glutamate uptake. *Nature* **348**, 443–447 (1990).
9. Kimelberg, H. K., Goderie, S. K., Higman, S., Pang, S. & Waniewski, R. A. Swelling-induced release of glutamate, aspartate and taurine from astrocyte cultures. *J. Neurosci.* **10**, 1583–1591 (1990).
10. Vesce, S., Bezzi, P., Rossi, D., Meldolesi, J. & Volterra, A. HIV-1 gp120 glycoprotein affects the astrocyte control of extracellular glutamate by both inhibiting the uptake and stimulating the release of the amino acid. *FEBS Lett.* **411**, 107–109 (1997).
11. Volterra, A. *et al.* The competitive transport inhibitor L-trans-pyrrolidine-2,4-dicarboxylate triggers excitotoxicity in rat cortical neuron-astrocyte co-cultures via glutamate release rather than uptake inhibition. *Eur. J. Neurosci.* **8**, 2019–2028 (1996).
12. Eriksson, P. S., Nilsson, M., Wägberg, M., Rönnebeck, L. & Hansson, E. Volume regulation of single astroglial cells in primary culture. *Neurosci. Lett.* **143**, 195–199 (1992).
13. Jętrinińska, S. D., Jętriniński, K. V. & Stefanovic, G. Cultured astrocytes express vesicles involved in vesicular glutamate release. *Brain Res.* **750**, 41–47 (1997).
14. Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992).
15. Jętrinińska, S. D., Jętriniński, K. V., Stefanovic, G. & Liu, F. Neurotrophin-induced calcium-dependent release of excitatory amino acids from cultured astrocytes. *J. Neurochem.* **66**, 674–684 (1996).
16. Dumuis, A., Pin, J. P., Oomagari, K., Sebben, M. & Bockaert, J. Arachidonic acid released from striatal neurons by joint stimulation of ionotropic and metabotropic quisqualate receptors. *Nature* **347**, 182–184 (1990).
17. Pimelli, D. Eicosanoids in synaptic transmission. *Crit. Rev. Neurobiol.* **8**, 65–83 (1994).
18. Oomagari, K., Buisson, B., Dumuis, A., Bockaert, J. & Pin, J. P. Effect of glutamate and ionomycin on the release of arachidonic acid, prostaglandins and HETEs from cultured neurons and astrocytes. *Eur. J. Neurosci.* **3**, 928–939 (1991).
19. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).
20. Calabresi, P., Benedetti, M., Mercuri, N. B. & Bernardi, G. Selective depression of synaptic transmission by tetanus toxin: a comparative study on hippocampal and neostriatal slices. *Neuroscience* **30**, 663–670 (1989).
21. Mennierick, S., Benz, A. & Zorumski, C. F. Components of glial responses to exogenous and synaptic glutamate in rat hippocampal microcultures. *J. Neurosci.* **16**, 55–64 (1996).
22. Pasti, L., Volterra, A., Pozzan, T. & Carmignoto, G. Intracellular calcium oscillations in astrocytes: a highly plastic, bidirectional form of communication between neurons and astrocytes *in situ*. *J. Neurosci.* **17**, 7817–7830 (1997).
23. Baran, H., Heldt, R. & Hertz, G. Increased prostaglandin formation in rat brain following systemic application of kainic acid. *Brain Res.* **404**, 107–112 (1987).
24. Malmberg, A. B. & Yaksh, T. L. Hyperalgesia mediated by spinal glutamate or substance P receptor blocked by spinal cyclooxygenase inhibition. *Science* **257**, 1276–1279 (1992).
25. Collaco-Moraes, Y., Asprey, B., Harrison, M. & de Belleruche, J. Cyclo-oxygenase-2 messenger RNA induction in focal cerebral ischemia. *J. Cereb. Blood Flow Metab.* **16**, 1366–1372 (1996).
26. O'Banion, M. K., Miller, J. C., Chang, J. W., Kaplan, M. D. & Coleman, P. D. Interleukin-1 β induces prostaglandin G/H synthase 2 (cyclooxygenase-2) in primary murine astrocyte cultures. *J. Neurochem.* **66**, 2532–2540 (1996).
27. Carmignoto, G. & Vicini, S. Activity-dependent decrease in NMDA receptor responses during development of the visual cortex. *Science* **258**, 1007–1011 (1992).
28. Stella, N., Tencé, M., Glowinski, J. & Prémont, J. Glutamate-evoked release of arachidonic acid from mouse brain astrocytes. *J. Neurosci.* **14**, 568–575 (1994).
29. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. A thin slice preparation for patch clamp recordings from synaptically connected neurons of the mammalian central nervous system. *Pflügers Arch.* **414**, 600–612 (1989).

Acknowledgements. We thank C. Montecucco for the generous supply of purified tetanus neurotoxin; J. Meldolesi and R. Paoletti for critical reading of the manuscript and advice; S. Nicosia for suggestions and use of facilities; and B. Viviani, M. R. Accomazzo and P. Ciceri for experimental help. This work was supported by grants from the European Community, 'Biomed 2 Contract BMH4-CT95-0571' and Telethon-Italy to A.V., and from Human Frontier Science Program RG520/95 and Telethon-Italy to T.P.

Correspondence and requests for materials should be addressed to A.V. (e-mail: Andrea.Volterra@unimi.it).

Induction of epithelial tubules by growth factor HGF depends on the STAT pathway

Carla Boccaccio, Margherita Andò, Luca Tamagnone, Alberto Bardelli, Paolo Michieli, Carlo Battistini* & Paolo M. Comoglio

Institute for Cancer Research, University of Torino Medical School, Strada Provinciale 142, 10060 Candiolo, Italy

*Pharmacia & Upjohn, Preclinical Research, 20014 Nerviano, Italy

Hepatocyte growth factor (HGF) induces a three-phase response leading to the formation of branched tubular structures in epithelial cells^{1,2}. The HGF receptor tyrosine kinase works through a Src homology (SH2) docking site that can activate several signalling pathways³. The first phase of the response (scattering), which results from cytoskeletal reorganization, loss of intercellular junctions and cell migration⁴, is dependent on phosphatidylinositol-3-OH kinase and Rac activation^{5,6}. The

second phase (growth) requires stimulation of the Ras–MAP kinase cascade⁷. Here we show that the third phase (tubulogenesis) is dependent on the STAT pathway. HGF stimulates recruitment of Stat-3 to the receptor, tyrosine phosphorylation, nuclear translocation and binding to the specific promoter element SIE. Electroporation of a tyrosine-phosphorylated peptide, which interferes with both the association of STAT to the receptor and STAT dimerization, inhibits tubule formation *in vitro* without affecting either HGF-induced 'scattering' or growth. The same result is obtained using a specific 'decoy' oligonucleotide that prevents STAT from binding to DNA and affecting the expression of genes involved in cell-cycle regulation (*c-fos* and *waf-1*). Activation of signal transducers that directly control transcription is therefore required for morphogenesis.

Various intracellular signalling pathways have been shown to be activated by receptor tyrosine kinases⁸. The precise role of each of these pathways must be resolved to understand the molecular basis of cell signalling specificity. The three biological responses mediated by the HGF receptor (scatter, growth and branching morphogenesis) are triggered by the tyrosine phosphorylation of a single multifunctional docking site located in the receptor's carboxy-terminal tail³. This sequence, containing two phosphotyrosines (Y¹³⁴⁹VHVNATY¹³⁵⁶VNV), interacts with several cytoplasmic signal transducers either directly or indirectly through molecular adaptors such as Grb2 (ref. 7), Shc (ref. 9) and Gab1 (ref. 10). We have previously shown that, after HGF stimulation, the receptor binds and activates phosphatidylinositol-3-OH kinase (PI(3)K)^{11,12} and recruits the Grb2–SoS complex, stimulating Ras¹³. The PI(3)K pathway is responsible for 'scattering' through the activation of Rac, and this in turn stimulates actin reorganization and other cytoskeletal changes that are required for cell motility^{5,6}. Mutagenesis of Asn 1358 in the docking site (necessary for recruiting the Grb2–SoS complex) generates an HGF receptor that can activate PI(3)K but not Ras. This mutant elicits cell scattering but not proliferation⁷. The multifunctional docking site is also directly responsible for branching morphogenesis. Insertion of the Y¹³⁴⁹VHVNATY¹³⁵⁶VNV sequence into the nerve growth factor (NGF) receptor, which otherwise lacks this biological response, makes it competent for induction of epithelial tubules². However, the signalling pathway(s) critical for tubulogenesis are not yet known. Here we show that the HGF receptor binds and phosphorylates Stat-3, and that the ensuing nuclear signalling is required to trigger differentiation for branching morphogenesis.

Association of STAT proteins with non-kinase receptors, such as those for interferons and cytokines, results in their phosphorylation on a conserved tyrosine residue by the cytoplasmic Jak kinases. This phosphorylation promotes the dimerization of STAT proteins through intermolecular interactions between phosphotyrosines and SH2 domains. STAT dimers translocate into the nucleus, where they act as transcriptional factors binding specific promoter sequences^{14,15}. We found that after HGF stimulation Stat-3 was transiently phosphorylated on tyrosine, with a peak between 15 and 30 min. Phosphorylated STAT was found predominantly in the nuclear fraction (Fig. 1 a–c), and its relocalization following HGF treatment was visualized by immunofluorescence (Fig. 1 d, e). STAT phosphorylation is likely to be mediated by the transient association of Stat-3 (through its SH2 domain) with the tyrosine-phosphorylated HGF receptor tail. This hypothesis is supported by three observations: first, co-precipitation of Stat-3 and the tyrosine-phosphorylated HGF receptor from intact cells (Fig. 2a); second, *in vitro* binding of endogenous Stat-3 with the purified recombinant cytoplasmic domain of the receptor (Fig. 2b); and third, binding of Stat-3 to the phosphorylated synthetic peptide Y^PVNV, corresponding to the sequence following Y¹³⁵⁶ in the HGF-receptor docking site (Fig. 2c). The interaction between Stat-3 and the receptor is both direct and indirect, through the adaptor Gab1, as shown by co-precipitation of this molecule with both Stat-3 (Fig. 2a) and the

HGF receptor¹⁰. The morphogenic response to HGF has been shown to be increased by transfection of Gab1 (ref. 10). The specificity of the interactions between Stat-3 and the receptor was demonstrated by competition experiments with the phosphopeptides Y^PVNV (which mimics the receptor docking sequence) and GY^PIKTE (which mimics the main STAT phosphorylation site, specific for binding to the STAT SH2 domain) (Fig. 2b).

Whether the STAT proteins are phosphorylated directly by tyrosine kinase receptors¹⁶ or indirectly through JAK activation¹⁷ is unclear. We found no evidence for complexes between JAK and the HGF receptor. Moreover, we found that the isolated HGF-receptor kinase domain phosphorylates *in vitro* a Stat-3 fragment (amino acids 580–715) containing the major phosphorylation site Y⁷⁰⁵ (data not shown).

STAT proteins are transcriptional factors that bind specific DNA elements containing highly conserved anti-palindromic sequences. In particular, STATs have high affinity for a modified form of the so-called 'sis inducible element' (h-SIE), which is part of the *c-fos* gene

promoter¹⁸. Mobility-shift analysis showed that HGF stimulation induces the appearance of nuclear complexes which specifically bind the SIE element. Formation of these complexes was prevented by adding the GY^PIKTE phosphopeptide to nuclear extracts (Fig. 3a).

In mice, targeted disruption of Stat-1 results in impaired interferon-mediated responses¹⁹ and disruption of Stat-5a in mammary-gland defects²⁰, whereas disruption of Stat-3 leads to embryonic lethality, suggesting that it is involved in early morphogenic processes²¹. To assess directly the role of Stat-3 in the three-step response elicited by HGF in epithelial cells (migration, growth and morphogenesis), we studied the effects of inhibiting the STAT signalling cascade at either end. On one side, we uncoupled the interaction of STAT with the HGF receptor (as well as STAT dimerization) by electroporation of the synthetic phosphopeptide GY^PIKTE. This phosphopeptide *in vitro* inhibits Stat-3 binding to the HGF-receptor docking site (Fig. 2b), prevents STAT dimerization¹⁴, and inhibits STAT binding to DNA (Fig. 3a). When electroporated in live cells, GY^PIKTE fully inhibits STAT translocation and binding to DNA (data not shown). Scattering and growth were unaffected (Fig. 4a,b), but formation of branched tubules was selectively inhibited (Fig. 4c). Specificity was demonstrated by testing the non-phosphorylated peptide and unrelated phosphopeptides, including that following Tyr 1313 (Y^PEVM), in the HGF-receptor tail. The Y^PVNV peptide, which uncouples all transducers from the receptor³, abolished all three responses (Fig. 4). At the opposite end of the signalling cascade, we inhibited binding of STAT to its specific DNA response elements by electroporating live cells with the 'decoy' oligonucleotide 5'-TTCCCGTAA-3', matching the conserved sequence of h-SIE. A scrambled oligonucleotide (5'-CTTGCCAAT-3') was used as a control. Inhibition of STAT binding to DNA once again left scatter and growth unaffected but abolished branching morphogenesis (data not shown).

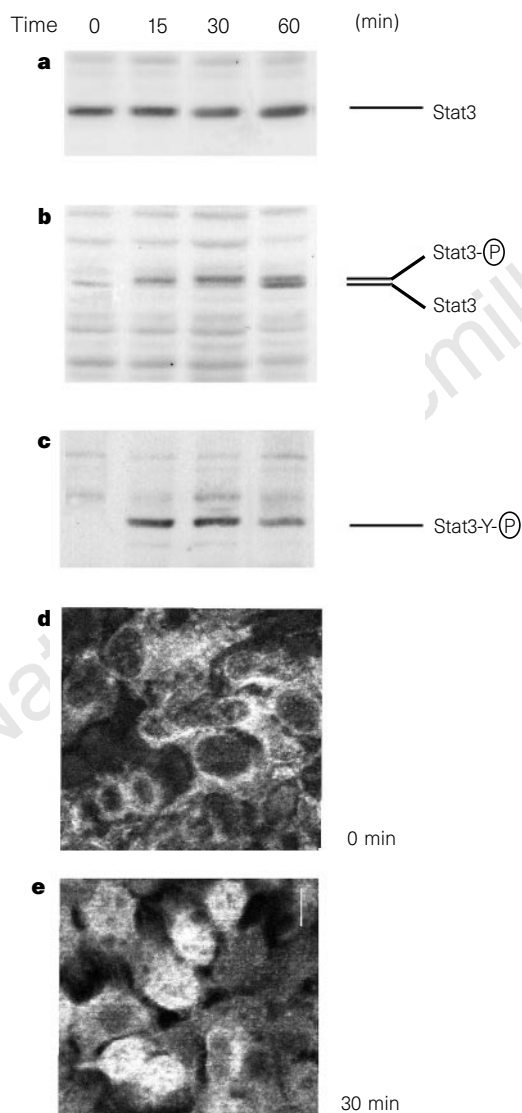


Figure 1 HGF induces STAT phosphorylation and translocation into the nucleus. **a–c**, MDCK cells were treated with HGF and extracts prepared from the cytoplasmic (**a**) or the nuclear (**b**, **c**) fractions. Western blot with anti-Stat-3 antibody detects a single band in the cytoplasm; a retarded extra band is detected in the nuclear extracts, corresponding to phosphorylated Stat-3, as assessed by western blot with anti-phosphotyrosine antibodies of specific immunoprecipitates (**c**). **d**, **e**, Phosphorylation is accompanied by translocation of Stat-3 from the cytoplasm (**d**) into the nucleus (**e**). Confocal microscopy: vertical scale bar, 10 μm.

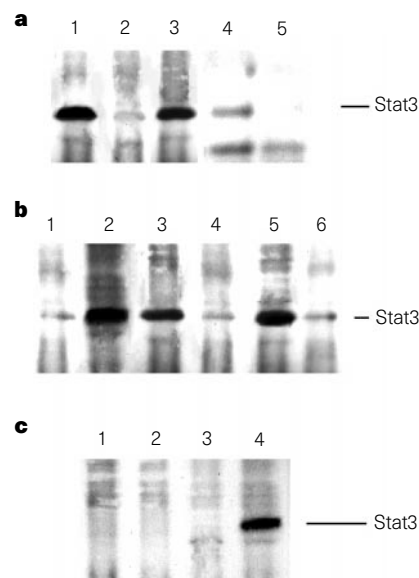


Figure 2 Stat-3 associates with the HGF receptor. **a**, Extracts of GTL-16 cells (featuring a constitutively tyrosine-phosphorylated HGF receptor) were immunoprecipitated with anti-HGF receptor (lane 1), anti-Stat-3 (lane 3), anti-Gab1 (lane 4) antibodies or non-immune mouse immunoglobulin (lanes 2 and 5). Immunocomplexes were western blotted with anti-Stat-3 antibodies. **b**, Cell lysates were pre-incubated with either buffer (lane 2) or 1 μM peptide YVNV (lane 3), Y^PVNV (lane 4), GYIKTE (lane 5) or GY^PIKTE (lane 6) and incubated with the tyrosine-phosphorylated HGF-receptor intracellular domain, coupled to Sepharose. Eluted proteins were analysed with anti-Stat-3 antibodies. Lane 1 shows the eluate from unphosphorylated receptor. **c**, Cell lysates were incubated with a resin coupled with Y^PEVM (lane 1), Y^PVHV (lane 2), YVNV (lane 3) and Y^PVNV (lane 4). Eluted proteins were analysed with anti-Stat-3 antibodies.

We therefore conclude that, in epithelial cells, the STAT signalling pathway is not required for short-term responses, including cell motility and growth, but is an absolute requirement to trigger long-term differentiation. Formation of branched tubular structures is driven by genes instructing the cells to stop growing and to differentiate, including *c-fos* and *waf-1* (a cell-cycle inhibitor), which contain specific STAT-binding elements in their promoters^{18,22}. *In vitro*, HGF stimulates transcription driven by the *c-fos* and *waf-1* promoters²³ (Fig. 3b,c). *In vivo*, HGF modulates the expression of the encoded proteins through the STAT pathway, as shown by interference of the specific 'decoy' oligonucleotide mimicking the h-SIE sequence (Fig. 3b,c). The decoy enhanced and prolonged the expression of the *c-fos* protein and inhibited p21^{Waf-1}

induction. The decoy-induced enhancement of *fos* expression is probably due to the quenching of the *fos* self-inhibition mechanism²⁴.

However, the STAT pathway, albeit necessary, is not sufficient for tubulogenesis. Other tyrosine kinases, such as the EGF receptor, are coupled to the same STAT proteins¹⁶ but are unable to induce branching morphogenesis^{1,2}. The response to HGF includes activation of other signalling molecules^{3,6,13}, so there is likely to be cross-talk between different transduction pathways. STAT binding to DNA is positively affected by serine phosphorylation at a C-terminal site which is a good consensus for MAP kinase¹⁵. The integration of the STAT pathway with other intracellular signals (through the conserved docking site which concomitantly activates multiple effectors) may explain why the HGF receptor family has the unique ability to promote branching morphogenesis. □

Methods

Biochemical and biological responses to HGF. Epithelial cell cultures (MDCK (ref. 4), GTL 16 (ref. 25) and MLP29 (ref. 1)) were grown and stimulated by 50 ng ml⁻¹ pure recombinant HGF as described³. Tyrosine phosphorylation and co-precipitation assays from subcellular fractions were performed. The antibodies used were a monoclonal anti-Stat-3 (Transduction), a monoclonal anti-HGF receptor β -chain³, and an anti-serum against Gab1 (Upstate Biotechnology). Binding to the intracellular domain of the HGF receptor was performed using a purified glutathione S-transferase (GST) fusion protein (amino acids 976–1408), expressed in baculovirus, immobilized on glutathione-Sepharose and autophosphorylated *in vitro* with excess ATP. Binding to peptides mimicking the HGF-receptor docking site was performed after coupling to 2 mg ml⁻¹ Actigel ALD resin (Sterogene, CA). Nuclear translocation was observed in cells fixed with 3% formaldehyde, permeabilized by 0.5% Triton and stained with anti-Stat-3 antibodies by indirect immunofluorescence. Electrophoretic mobility-shift assays used a ³²P-dCTP radio-

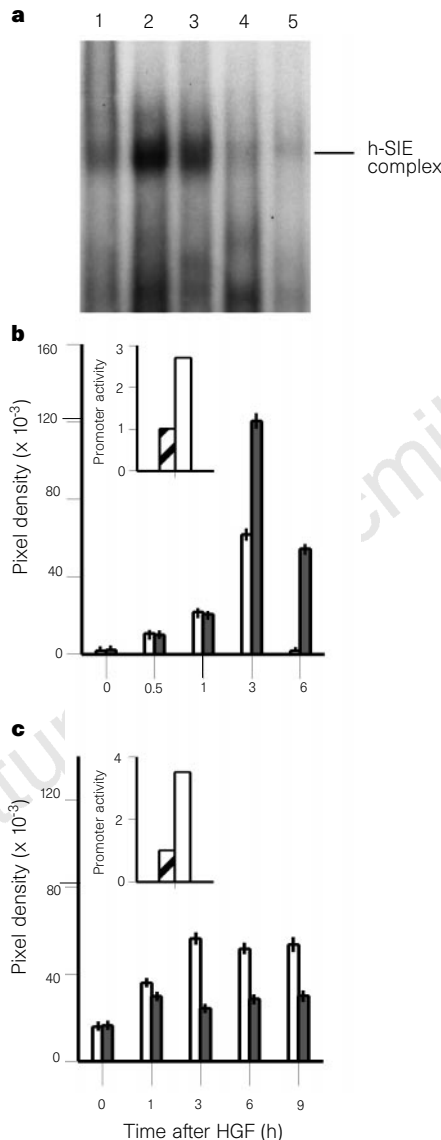


Figure 3 The SIE elements binds STAT and modulates the expression of HGF-induced genes. **a**, Electromobility-shift assay of nuclear extracts from MDCK cells challenged with the h-SIE probe¹⁸. Unstimulated cells (lane 1), cells treated with HGF for 15 min (lane 2), as above but with preincubation of lysates with 1 μ M GYIKTE (lane 3), GY^PIKTE (lane 4) or unlabelled h-SIE probe (lane 5). **b**, **c**, Cells were electroporated with h-SIE probe (black bars) or with a control scrambled sequence (white bars) and stimulated with HGF. **b**, p62^{c-fos} was quantified in western blots by densitometric analysis. The inset shows the *c-fos* promoter-driven luciferase transcription after stimulation by HGF (white bar) or buffer (hatched bar). **c**, The p21^{Waf-1} response, and the stimulation of the p21^{Waf-1} promoter (inset).

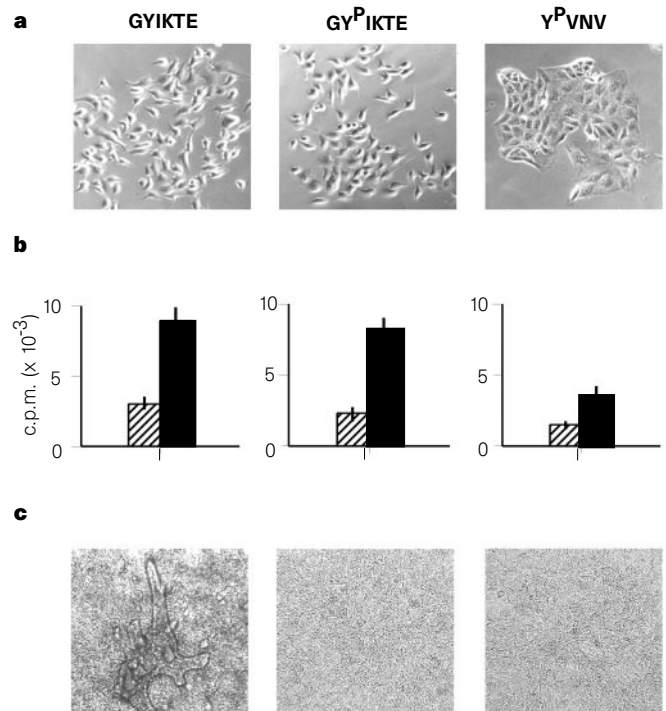


Figure 4 Selective inhibition of branching morphogenesis by a STAT phosphopeptide. Cells were electroporated *in situ* with the peptides indicated above each column; the phosphate group was substituted with methylenphosphonate, resistant to intracellular tyrosine phosphatases³⁰. **a**, Micrographs ($\times 1,000$) of the 'scattering' response 6 h after HGF stimulation. **b**, Histograms of ³H-thymidine incorporation in the absence (hatched bars) or the presence of HGF (black bars). **c**, Micrographs ($\times 40$) of monolayers coated with a three-dimensional array of collagen and incubated for 3 days in the presence of HGF.

labelled h-SIE probe (5'-CATTTCCCGTAAATC-3')¹⁸ containing a protruding poly-G tail at both sense and antisense 5' ends. HGF-induced stimulation of transcriptional activities of the isolated *c-fos* (ref. 26) and *waf-1* (ref. 27) promoters was measured by the Luciferase assay²³. Synthesis of p62^{c-fos} and p21^{waf-1} was measured by western blot with specific antibodies (gifts from C. Schneider). Cell 'scattering' in response to HGF (50 ng ml⁻¹) was measured as described²⁸. Micrographs (×1,000) were taken 6 h after stimulation. Stimulation of growth was assessed by ³H-thymidine uptake 12 h after HGF stimulation of cells synchronized before electroporation by two cycles of thymidine (2 mM)-deoxycytidine (250 μM). Formation of branched tubular structures was monitored on cell monolayers, coated with a 0.5-mm layer of collagen G (2 mg ml⁻¹) in the presence of HGF (10 ng ml⁻¹) supplemented daily¹.

Phosphopeptides, decoys and electroporation. N-acetylated and C-amidated synthetic peptides, mimicking the consensus sequences for SH2-signal transducers in the HGF receptor or in the STAT family, were synthesized in their phosphorylated or non-phosphorylated forms by Fmoc chemistry, and purified by high-performance liquid chromatography (HPLC; >99%), as will be described elsewhere. For *in vivo* electroporation the phosphate group was substituted by the non-hydrolysable methylenphosphonate group. The h-SIE decoy (5'-CATTTCCCGTAAATC-3') and its scrambled control 5'-ACTCTTGCCAATTAC-3' were synthesized and used in the inhibition assay as described²⁸. For electroporation, adherent cells were cultured onto 1-cm² indium-tin oxide conductive glass surfaces and electroporated *in situ* with a PBS solution containing either the competing peptides (1 μM) or the decoys (10 μM). The electric pulse was generated using the Epizap apparatus (Ask Science, Canada), following the technique described previously²⁹. More than 90% of cells were permeabilized as tested by the uptake of Lucifer yellow, and viability was higher than 80% as judged by trypan blue exclusion.

Received 4 August; accepted 6 October 1997

- Medico, E. *et al.* The tyrosine kinase receptors Ron and Sea control "scattering" and morphogenesis of liver progenitor cells *in vitro*. *Mol. Biol. Cell* **7**, 495–504 (1996).
- Sachs, M. *et al.* Mitogenic and morphogenic activity of epithelial receptor tyrosine kinases. *J. Cell Biol.* **133**, 1095–1107 (1996).
- Ponzetto, C. *et al.* A multifunctional docking site mediates signalling and transformation by the Hepatocyte Growth Factor/Scatter Factor receptor family. *Cell* **77**, 261–271 (1994).
- Stocker, M., Gherardi, E., Perryman, M. & Gray, J. Scatter factor is a fibroblast-derived modulator of epithelial cell motility. *Nature* **327**, 239–242 (1987).
- Ridley, A. J., Comoglio, P. M. & Hall, A. Regulation of Scatter Factor/Hepatocyte Growth Factor responses by Ras, Rac, and Rho in MDCK cells. *J. Cell Biol.* **15**, 1110–1122 (1995).
- Royal, I. & Park, M. Hepatocyte Growth Factor-induced scatter of Madin–Darby canine kidney cells requires phosphatidylinositol 3-kinase. *J. Biol. Chem.* **270**, 27780–27787 (1995).
- Ponzetto, C. *et al.* Specific uncoupling of GRB2 from the Met receptor. *J. Biol. Chem.* **271**, 14119–14123 (1996).
- Pawson, T. Protein modules and signalling networks. *Nature* **373**, 573–580 (1995).
- Pellici, G. *et al.* The mitogenic and mitogenic responses to HGF are amplified by the Shc adaptor protein. *Oncogene* **10**, 1631–1638 (1995).
- Weidner, K. M. *et al.* Interaction between Gab1 and the c-Met receptor tyrosine kinase is responsible for epithelial morphogenesis. *Nature* **384**, 173–176 (1996).
- Graziani, A., Gramaglia, D., Cantley, L. C. & Comoglio, P. M. The tyrosine-phosphorylated hepatocyte growth factor/scatter factor receptor associates with phosphatidylinositol 3-kinase. *J. Biol. Chem.* **266**, 22087–22090 (1991).
- Ponzetto, C. *et al.* A novel recognition motif for Phosphatidylinositol 3-kinase binding mediates its association with the Hepatocyte Growth Factor/Scatter Factor receptor. *Mol. Cell Biol.* **13**, 4600–4608 (1993).
- Graziani, A., Gramaglia, D., dalla Zonca, P. & Comoglio, P. M. Hepatocyte Growth Factor/Scatter Factor stimulates the Ras-guanine nucleotide exchanger. *J. Biol. Chem.* **268**, 9165–9168 (1993).
- Schindler, C. & Darnell, J. E. Jr Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annu. Rev. Biochem.* **64**, 621–651 (1995).
- Ilhe, J. N. STATs: signal transducers and activators of transcription. *Cell* **84**, 221–334 (1996).
- Leaman, D. W. *et al.* Roles of JAKs in activation of STATs and stimulation of *c-fos* gene expression by Epidermal Growth Factor. *Mol. Cell Biol.* **16**, 369–375 (1996).
- Silvennoinen, O., Ilhe, J. N., Schlessinger, J. & Levy, D. E. Interferon-induced nuclear signalling by Jak protein tyrosine kinases. *Nature* **366**, 583–585 (1993).
- Wagner, B. J., Hayes, T. E., Hoban, C. J. & Cochran, B. H. The SIF binding element confers sis/PDGF inducibility onto the *c-fos* promoter. *EMBO J.* **9**, 4477–4484 (1990).
- Durbin, J. E., Hackenmiller, R., Simon, M. C. & Levy, D. E. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell* **84**, 443–450 (1996).
- Liu, X. *et al.* Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* **11**, 179–186 (1997).
- Takeda, K. *et al.* Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc. Natl Acad. Sci. USA* **94**, 3801–3804 (1997).
- Chin, Y. E. *et al.* Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21^{WAF1/CIP1} mediated by STAT1. *Science* **272**, 719–722 (1996).
- Boccaccio, C., Gaudino, G., Gambardella, G., Galimi, F. & Comoglio, P. M. Hepatocyte Growth factor (HGF) receptor expression is inducible and is part of the delayed-early response to HGF. *J. Biol. Chem.* **269**, 12846–12851 (1994).
- Lucibello, F. C., Lowag, C., Neuber, M. & Muller, R. Trans-repression of the mouse *c-fos* promoter: a novel mechanism of Fos-mediated trans-regulation. *Cell* **59**, 999–1007 (1989).
- Giordano, S., Ponzetto, C., Di Renzo, M. E., Cooper, C. S. & Comoglio, P. M. Tyrosine kinase receptor indistinguishable from the c-Met protein. *Nature* **339**, 155–156 (1989).
- Gilman, M. Z., Wilson, R. N. & Weinberg, R. A. Multiple protein binding sites in the 5'-flanking region regulate *c-fos* expression. *Mol. Cell Biol.* **6**, 4305–4316 (1986).

- El-Deiry, W. S. *et al.* WAF-1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Gambardella, G. *et al.* Ets up-regulates MET transcription. *Oncogene* **13**, 1911–1917 (1996).
- Raptis, L. H., Liu, S. K. W., Firth, K. L., Stiles, C. D. & Alberta, J. A. Electroporation of peptides into adherent cells *in situ*. *Biotechniques* **19**, 104–114 (1995).
- Comoglio, P. M. *et al.* Detection of phosphotyrosine containing proteins in the detergent insoluble fraction of RSV-transformed fibroblasts by azobenzene phosphonate antibodies. *EMBO J.* **3**, 483–489 (1984).

Acknowledgements. We thank P. Giordano (Pharmacia & Upjohn) for peptide synthesis; C. Ponzetto for discussions; A. Cignetto for secretarial help; and E. Wright for help with the manuscript. This work was supported by an AIRC grant to P.M.C. M.A. is recipient of a FIRC Fellowship.

Correspondence and requests for materials should be addressed to C. Boccaccio (e-mail: cboccaccio@hal.ircc.polito.it).

DNA shuffling of a family of genes from diverse species accelerates directed evolution

Andreas Crameri, Sun-Ai Raillard, Ericka Bermudez & Willem P. C. Stemmer

Maxygen Inc., 3410 Central Expressway, Santa Clara, California 95051, USA

DNA shuffling is a powerful process for directed evolution, which generates diversity by recombination^{1,2}, combining useful mutations from individual genes. Libraries of chimaeric genes can be generated by random fragmentation of a pool of related genes, followed by reassembly of the fragments in a self-priming polymerase reaction. Template switching causes crossovers in areas of sequence homology. Our previous studies used single genes and random point mutations as the source of diversity^{3–6}. An alternative source of diversity is naturally occurring homologous genes, which provide 'functional diversity'. To evaluate whether natural diversity could accelerate the evolution process, we compared the efficiency of obtaining moxalactamase activity from four cephalosporinase genes evolved separately with that from a mixed pool of the four genes. A single cycle of shuffling yielded eightfold improvements from the four separately evolved genes, versus a 270- to 540-fold improvement from the four genes shuffled together, a 50-fold increase per cycle of shuffling. The best clone contained eight segments from three of the four genes as well as 33 amino-acid point mutations. Molecular breeding by shuffling can efficiently mix sequences from different species, unlike traditional breeding techniques. The power of family shuffling may arise from sparse sampling of a larger portion of sequence space.

Iterative cycles of shuffling followed by screening or selection has proved to be a useful approach for the evolution of single gene products with enhanced activity³, altered substrate specificity⁴ or improved protein folding⁵ and of entire operons with improved function⁶. When a single starting sequence is used, diversity originates as random point mutations resulting from the polymerase reaction¹. Because most point mutations are deleterious or neutral⁷, the random point mutation rate must be low⁸ and the accumulation of beneficial mutations and the evolution of a desired function is relatively slow in such experiments. For example, the evolution of a fucosidase from a galactosidase required five rounds of shuffling and screening before a >10-fold improvement in activity was detected⁴. Naturally occurring homologous sequences are pre-enriched for 'functional diversity' because deleterious variants

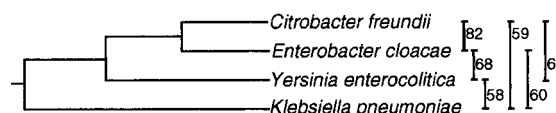


Figure 1 Phylogenetic tree of the four cephalosporinase genes. The numbers on the vertical bars indicate the percentage of DNA sequence similarity.

have been selected against over billions of years of evolution. We therefore wanted to determine whether shuffling of gene families would accelerate the evolution process.

Four 1.6 kilobase (kb) genes encoding class C cephalosporinases, 58–82% identical at the DNA sequence level (Fig. 1), were chosen from four microbial species (*Citrobacter freundii*⁹, *Enterobacter cloacae*¹⁰, *Klebsiella pneumoniae*¹¹ and *Yersinia enterocolitica*¹²) and constructed by gene synthesis¹³ using the original DNA sequence. The four genes were shuffled either individually or as a pool using methods previously described^{1,3} (Fig. 2a). Equal numbers of *Escherichia coli* transformants (about 5×10^4) expressing the resultant libraries were then plated on media containing a range of concentrations of the antibiotic moxalactam (0.016 – $32 \mu\text{g ml}^{-1}$). After this single round of shuffling, the colonies showing the highest level of resistance to moxalactam were identified and further characterized.

Clones originating from the four single gene libraries showed up to eightfold increases in moxalactam resistance as compared to those expressing the wild-type genes (0.38 to $3.0 \mu\text{g ml}^{-1}$ for the *Klebsiella* and *Yersinia* genes and 0.75 to $6.0 \mu\text{g ml}^{-1}$ for the *Citrobacter* and *Enterobacter* genes). In contrast, the best clone (A) originating from the gene family library showed a 540-fold increase in resistance (0.38 to $200 \mu\text{g ml}^{-1}$) compared to the wild-type *Klebsiella* and *Yersinia* genes and a 270-fold increase (0.75 to $200 \mu\text{g ml}^{-1}$) compared to the wild-type *Enterobacter* and *Citrobacter* genes (Fig. 2a). Thus 'family shuffling' accelerated the rate of functional enzyme improvement 34- to 68-fold in a single cycle.

A second round of shuffling was performed using the pool of colonies selected in the first round. Plating of about 5×10^4 colonies on a range of concentrations of moxalactam yielded three clones which had a further 3.5-fold increased moxalactam resistance over the most resistant clone for the first cycle (clone A). The reduction in the rate of improvement is expected given the limited dynamic range of the bioassay. Based on our experience with other genes^{3–6}, a second round of intra-species shuffling of the

selected pools containing the eightfold improved genes would clearly not have achieved the degree of improvement we obtained by two rounds of inter-species shuffling.

The two most resistant clones (A and B) obtained in the first round of family shuffling also showed increased resistance against other β -lactam antibiotics. Clone A was resistant to cefoxitin, carbenicillin and cephafloridine at $100 \mu\text{g ml}^{-1}$ each, an improvement over the four wild-type enzymes of 4–16-fold. This was unexpected because in previous results for single gene shuffling of other enzymes the activity for the original substrate was decreased after evolution for increased activity on a new substrate¹⁴. Plasmid transfer experiments into *E. coli* NM522 showed that all of the increased resistance was conferred by the plasmid. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis of periplasmic extracts showed that the expression level of the four wild-type cephalosporinases was indistinguishable from that of the mutant clones A and B (data not shown). Antibiotic bioassays showed that all of the moxalactam was degraded by cells expressing chimaeric enzyme A or B, but not by cells expressing any of the four wild-type enzymes. This indicates that the specific activity of the chimaeric moxalactamases was improved. However, our attempts to obtain and compare the kinetic profiles of the wild-type and chimaeric enzymes failed because of the very low activity of the wild-type enzymes.

The genes encoding the new moxalactamases from these two clones were sequenced. Clones A and B were both found to be chimaeras of the genes from *Citrobacter*, *Enterobacter* and *Klebsiella*. Both clones had a similar overall structure containing eight segments resulting from seven crossovers. Because of local DNA homology the crossover location could not be defined more exactly than indicated by the grey segments shown in Fig. 2b. The crossovers occurred in areas where the two genes had 14–37 base pairs (bp) of nearly identical sequence (average 57% GC). In addition, chimaera A had 47 DNA point mutations, 6 of which were silent, resulting in 33 amino-acid substitutions scattered throughout the gene that did not exist in any of the four parental enzymes.

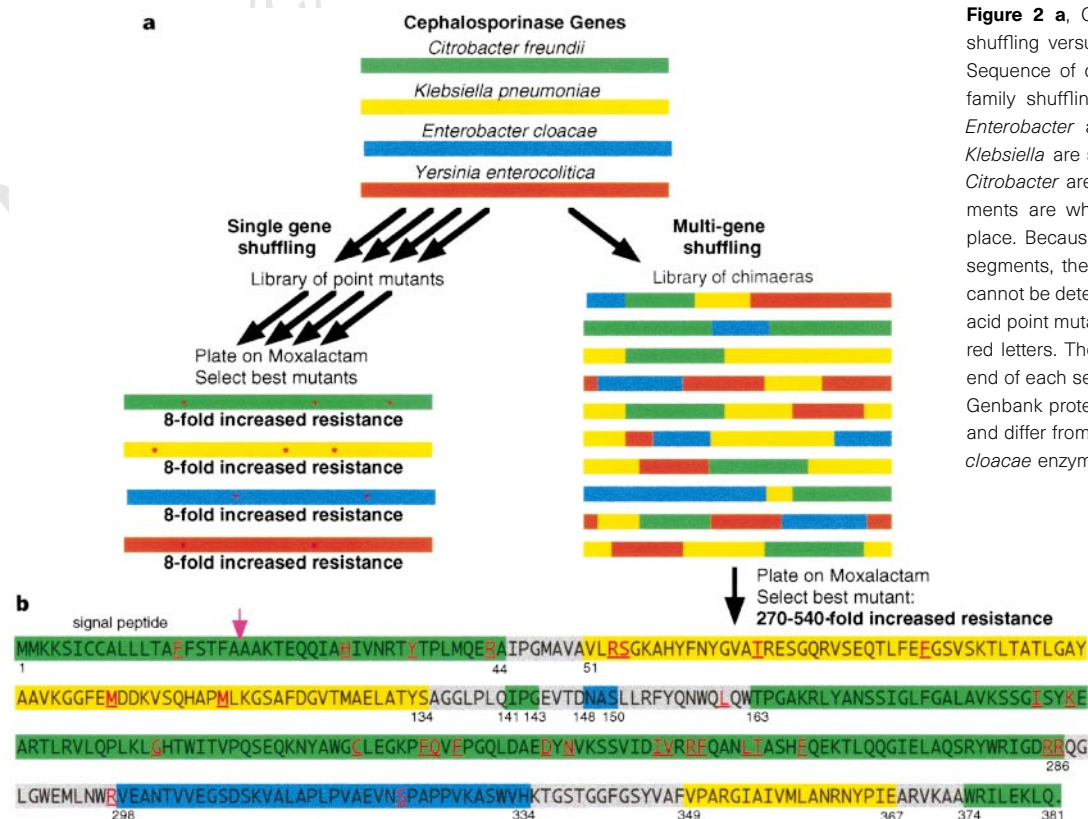


Figure 2 a, Comparison of single sequence shuffling versus sequence family shuffling. b, Sequence of chimaeric mutant A obtained by family shuffling. The segments derived from *Enterobacter* are shown in blue, those from *Klebsiella* are shown in yellow, and those from *Citrobacter* are shown in green. The grey segments are where the crossovers have taken place. Because of DNA homology in the grey segments, the exact location of the crossover cannot be determined more exactly. The amino-acid point mutations are shown with underlined red letters. The numbers at the beginning and end of each segment are the numbers from the Genbank protein files of the wild-type enzymes and differ from those used for the *Enterobacter cloacae* enzyme¹⁴.

Chimaera B had 14 amino-acid substitutions, of which 12 were identical to those in chimaera A.

A model of the best chimaeric moxalactamase, chimaera A, was created from the known structure of the *Enterobacter cloacae* AmpC enzyme¹⁴ (Fig. 3). Although 37% of the amino acids (142 residues) of the chimaeric clone A differ from the *Enterobacter cloacae* enzyme for which the crystal structure is known, after energy minimization the predicted structure of the α -chain backbone of the A chimaera remained nearly identical to the known structure (r.m.s. deviation of 0.766 Å). This model shows that two crossovers occurred in loop or random coil regions separating α -helical and β -sheet structures, and one occurred inside the C-terminal α -helix. The remaining five crossovers could have occurred in loops, α -helices or β -sheets; therefore it is unclear whether the crossovers preferentially occurred in loops separating structural elements. The conserved catalytic residues S64, K67,

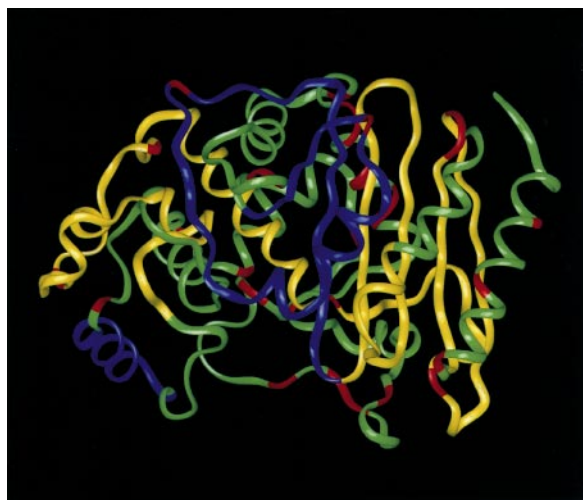


Figure 3 Computer model of evolved mutant A obtained by a single cycle of family shuffling. The 142 amino-acid mutations were introduced into the *Enterobacter cloacae* sequence, whose structure is known¹⁴, followed by energy minimization. The predicted structure of the α -chain backbone of the chimaeric enzyme is within an r.m.s. deviation of 0.766 Å from the known native structure. The segments derived from *Enterobacter* are shown in blue, those from *Klebsiella* are shown in yellow, and those from *Citrobacter* are shown in green. The 33 amino-acid point mutations are shown in red.

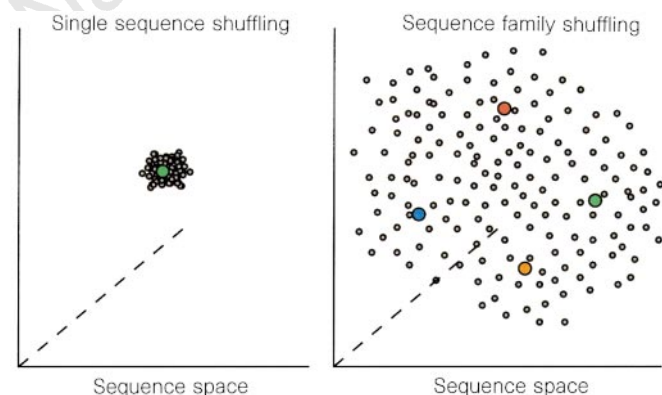


Figure 4 Searching sequence space by family shuffling versus by single sequence shuffling. Single sequence shuffling yields clones with a few point mutations and the library members are typically 97–99% identical. Family shuffling causes sequence block exchange which yields chimaeras that have greater sequence divergence. At equal library size, the increased sequence diversity of the chimaeric library results in sparse sampling of a much greater area of sequence space, allowing more promising areas to be found and subsequently explored at increased sampling density.

Y150 and K315 of the *Enterobacter* enzyme¹⁴ were retained.

Although shuffling of a single gene creates a library of genes that differ by only a few point mutations^{1–6}, the block-exchange nature of family shuffling creates chimaeras that differ in many positions. For example, in previous work a single β -lactamase gene was shuffled for three cycles, yielding only four amino-acid mutations³, whereas a single cycle of family shuffling of the four cephalosporinases resulted in a mutant enzyme which differs by 102 amino acids from the *Citrobacter* enzyme, by 142 amino acids from the *Enterobacter* enzyme, by 181 amino acid from the *Klebsiella* enzyme and by 196 amino acids from the *Yersinia* enzyme. The increased sequence diversity of the library members obtained by family shuffling results in a 'sparse sampling' of a much greater portion of sequence space¹⁵, the theoretical collection of all possible sequences of equal length, ordered by similarity (Fig. 4). Selection from 'sparse libraries' allows rapid identification of the most promising areas within an extended sequence landscape (a multidimensional graph of sequence space versus function)¹⁵. However, the sparseness also decreases the likelihood of immediate location of the area's best sequence, the local optimum. Subsequent shuffling of the selected sequences allows further exploration of the still vast intermittent sequence space at an increased sampling density. Although the search algorithm remains unchanged, the scale of the searched area decreases with each cycle until no further improvement occurs.

Because the species definition in eukaryotes is based on the ability to exchange genetic information, the family shuffling that occurs in natural evolution and in classical breeding is generally restricted to the diversity existing within a single species. DNA shuffling can successfully mix genes from diverse species because single genes can tolerate much higher mutation densities¹⁶ than whole genomes. The high degree of DNA homology that is required for traditional breeding of plants and animals is therefore not required for the molecular breeding of single genes and gene clusters. □

Methods

Gene synthesis. All four 1.6 kb genes were separately assembled from 60-mer synthetic oligonucleotides in a single assembly reaction¹³. These genes were cloned into pBR322 downstream of the β -lactamase promoter and in place of the native β -lactamase gene. The DNA sequence and drug resistance of each construct was confirmed.

Library construction. The presence of chimaeric genes in the library made by family shuffling was confirmed by restriction fragment length polymorphism (RFLP) analysis of individual clones. Each of the eight clones analysed had an RFLP pattern distinct from each other as well as from those of the parental sequences. The library thus contained a diverse array of chimaeric sequences. The second cycle of shuffling was performed on the pool of chimaeras selected in the first round.

Drug resistance. An equal number of transformants were plated on medium containing moxalactam. The numbers were based on the number of tetracyclin-resistant colonies after transformation. The resistance of the selected clones to a variety of β -lactam antibiotics was measured in 96-well plates in which each antibiotic was serially diluted (1 : 1) into culture medium containing the test organism. Assays for all native and mutant clones were done in quadruplicate.

Moxalactamase assays. The moxalactam degradation activity of whole cells was measured by incubating an overnight culture of bacteria with 34 $\mu\text{g ml}^{-1}$ of moxalactam in LB media. After incubation for 16 h the concentration of moxalactam in the sterilized supernatant was measured by bioassay on plates inoculated with *E. coli* NM522, using filter discs saturated with serial twofold dilutions of the supernatants. A moxalactam standard curve was used to determine the fraction of moxalactam degraded by the host cell alone, the four wild-type constructs and the two chimaeric mutants.

Modelling. The deduced amino-acid sequence of chimaeric mutant A was aligned with the wild-type *Enterobacter* sequence using HOMOLGY (Biosym, San Diego) and the calculated coordinates were assigned to the chimaeric protein. The modelled structure was constructed using a three-step protocol that involved an energy minimization to relieve steric hindrance, molecular

dynamics calculations to find the lowest energy conformation followed by another energy minimization to provide the final structure.

Received 23 June; accepted 8 October 1997.

1. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl Acad. Sci. USA* **91**, 10747–10751 (1994).
2. Stemmer, W. P. C. Searching sequence space. *BioTechnology* **13**, 549–553 (1995).
3. Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
4. Zhang, J., Dawes, G. & Stemmer, W. P. C. Evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl Acad. Sci. USA* **94**, 4504–4509 (1997).
5. Cramer, A., Whitehorn, E., Tate, E. & Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotech.* **14**, 315–319 (1996).
6. Cramer, A., Dawes, G., Rodriguez, E., Silver, S. & Stemmer, W. P. C. Molecular Evolution of an arsenate detoxification pathway by DNA shuffling. *Nature Biotech.* **15**, 436–438 (1997).
7. Moore, J. C. & Arnold, F. H. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotech.* **14**, 458–467 (1996).
8. Zhao, H. & Arnold, F. H. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307–1308 (1997).
9. Lindberg, F. & Normark, S. Sequence of the *Citrobacter freundii* OS60 chromosomal *ampC* β -lactamase gene. *Eur. J. Biochem.* **156**, 441–445 (1986).
10. Galleni, M. *et al.* Sequence and comparative analysis of three *Enterobacter cloacae* *ampC* β -lactamase genes and their products. *Biochem. J.* **250**, 753–760 (1988).
11. Leiza, M. G. *et al.* Gene sequence and biochemical characterization of FOX-1 from *Klebsiella pneumoniae*, a new *AmpC*-type plasmid-mediated β -lactamase with two molecular variants. *Vita-Antimicrob. Agents Chemother.* **38**, 2150–2157 (1994).
12. Seoane, A., Francia, M. V. & Garcia Lobo, J. M. Nucleotide sequence of the *ampC*-*ampR* region from the chromosome of *Yersinia enterocolitica*. *Antimicrob. Agents Chemother.* **36**, 1049–1052 (1992).
13. Stemmer, W. P. C., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step PCR assembly of a gene and a whole plasmid from large numbers of oligonucleotides. *Gene* **164**, 49–53 (1995).
14. Lobkovsky, E. *et al.* Evolution of enzyme activity: crystallographic structure at 2 Å resolution of cephalosporinase from the *ampC* gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc. Natl Acad. Sci. USA* **90**, 11257–11261 (1993).
15. Kauffman, S. *The Origins of Order* (Oxford University Press, Oxford, 1993).
16. Eigen, M. *Steps Towards Life: A Perspective on Evolution* (Oxford University Press, Oxford, 1992).

Acknowledgements. We thank G. Dawes, J. Kieft, S. DelCardayre and M. Tobin and R. Howard, C. Yanofsky, P. Schultz, F. Arnold and A. Kornberg for useful comments on the manuscript.

Correspondence should be addressed to W.P.C.S. (e-mail: pim_stemmer@maxygen.com).

A bacterial antibiotic-resistance gene that complements the human multidrug-resistance P-glycoprotein gene

Hendrik W. van Veen*, Richard Callaghan†, Loredana Soceneantu†, Alessandro Sardini‡, Wil N. Konings* & Christopher F. Higgins†

* Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

† Imperial Cancer Research Fund Laboratories and Cancer Research Campaign Drug Resistance Group, Nuffield Department of Clinical Biochemistry, Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK

‡ Department of Physiology, King's College London, The Strand, London WC2R 2LS, UK

Bacteria have developed many fascinating antibiotic-resistance mechanisms^{1,2}. A protein in *Lactococcus lactis*, LmrA, mediates antibiotic resistance by extruding amphiphilic compounds from the inner leaflet of the cytoplasmic membrane^{3,4}. Unlike other known bacterial multidrug-resistance proteins, LmrA is an ATP-binding cassette (ABC) transporter⁵. The human multidrug-resistance P-glycoprotein⁶, encoded by the *MDR1* gene, is also an ABC transporter, overexpression of which is one of the principal causes of resistance of human cancers to chemotherapy^{7,8}. We expressed *lmrA* in human lung fibroblast cells. Surprisingly, LmrA was targeted to the plasma membrane and conferred typical multidrug resistance on these human cells. The pharmacological characteristics of LmrA and P-glycoprotein-expressing lung fibroblasts were very similar, and the affinities of both proteins for

vinblastine and magnesium-ATP were indistinguishable. Blockers of P-glycoprotein-mediated multidrug resistance also inhibited LmrA-dependent drug resistance. Kinetic analysis of drug dissociation from LmrA expressed in plasma membranes of insect cells revealed the presence of two allosterically linked drug-binding sites indistinguishable from those of P-glycoprotein. These findings have implications for the reversal of antibiotic resistance in pathogenic microorganisms. Taken together, they demonstrate that bacterial LmrA and human P-glycoprotein are functionally interchangeable and that this type of multidrug-resistance efflux pump is conserved from bacteria to man.

Using the polymerase chain reaction (PCR), the bacterial *lmrA* coding sequence³ was cloned into a pCI-neo mammalian expression vector under the control of the human cytomegalovirus immediate-early enhancer/promoter region. A Kozak sequence⁹ was introduced at the ATG initiation codon of *lmrA* to enhance translational efficiency. For control experiments, a transport-inactive LmrA protein was generated in the same vector by introducing a lysine-to-methionine substitution at position 388 (K388M)¹⁰ in the Walker A motif of the nucleotide-binding domain of the protein by site-directed mutagenesis. Hexa-histidine tags were also added to the amino termini of both the wild-type and K388M forms of LmrA.

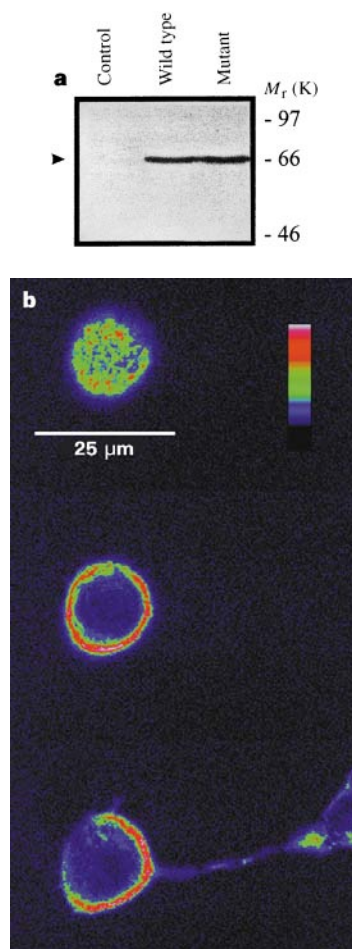


Figure 1 Expression of LmrA in GM0637 human lung fibroblast cells. **a**, Western blot of total cell protein (30 µg per lane) using the anti-hexa-histidine tag antibody. Control, wild type and mutant refer to mock-transfected cells, and cells expressing the wild-type and K388M forms of LmrA, respectively. The migration of molecular mass markers is indicated. Arrowhead indicates LmrA protein. **b**, Distribution of wild-type LmrA expressed in fibroblast cells 48 h after transfection. The three frames show cross-sections through a cell, from top to bottom along the z-axis. The fluorescence intensity is shown colour-coded on a scale from blue (low) to white (high). Scale bar, 25 µm.

LmrA was transiently expressed in the GM0637 lung fibroblast cell line that had previously been used for molecular characterization of P-glycoprotein¹¹. Transfection of both the wild-type and mutant *lmrA* genes into the fibroblast cells resulted in the expression of equivalent amounts of a 66K polypeptide, which could be detected by using an anti hexa-histidine tag monoclonal antibody (Fig. 1a). Confocal fluorescence microscopy indicated that, as for human P-glycoprotein⁷, LmrA was primarily localized in the plasma membrane of the fibroblast cells (Fig. 1b). Transfected GM0637 lung fibroblast cells contained undetectable levels of P-glycoprotein¹¹ or the multidrug-resistance-associated protein MRP1¹² (data not shown).

Lung fibroblast cells expressing wild-type LmrA protein showed a 10–60-fold increased resistance to a variety of natural product drugs and synthetic chemotherapeutic drugs which are typical substrates of P-glycoprotein⁷: (1) anthracyclines such as daunomycin and doxorubicin; (2) vinca-alkaloids such as vinblastine and vincristine; and (3) cytotoxic agents such as ethidium bromide, rhodamine 123 and colchicine (Fig. 2a, b). Mock-transfected control cells, and cells expressing equivalent amounts of the K388M mutant LmrA, showed no increase in resistance, demonstrating that LmrA itself is responsible for the drug-resistance profile.

P-glycoprotein function can be inhibited by a large group of structurally unrelated modulators⁷. All of these modulators also reversed drug resistance generated by expression of wild-type LmrA

in lung fibroblast cells: (1) calcium-channel blockers such as verapamil and its analogue CP100-356; (2) 1,4-dihydropyridines such as nicardipine; (3) indolizine sulphones such as SR33557; (4) antimalarials such as quinine and quinidine; and (5) immunosuppressants such as cyclosporin A (Fig. 2c, d).

To demonstrate that drug extrusion across the plasma membrane is the underlying mechanism of drug resistance in LmrA-expressing lung fibroblast cells, we did transport assays using both vinblastine and Hoechst 33342 (Fig. 3a, b). The accumulation of these compounds was significantly reduced in cells expressing wild-type LmrA. LmrA-mediated drug efflux was inhibited by the *Rauwolfia* alkaloid reserpine and by the phenylalkylamine verapamil, two classic inhibitors of drug transport by P-glycoprotein⁷. These inhibitors did not affect the accumulation of vinblastine or Hoechst 33342 in cells expressing the K388M mutant form of LmrA (data not shown).

To explore the pharmacology of wild-type LmrA and to compare it with that of P-glycoprotein, larger quantities of both proteins were expressed in recombinant baculovirus-infected *Spodoptera frugiperda* insect cells¹³ (data not shown). LmrA and P-glycoprotein displayed a basal ATPase activity which was stimulated up to three-fold by 50 μ M verapamil and completely inhibited by 2 mM vanadate, and which were comparably dependent on the concentration of magnesium-ATP (LmrA: half-maximal ATPase activity at 0.24 ± 0.04 mM Mg-ATP; P-glycoprotein: half-maximal ATPase

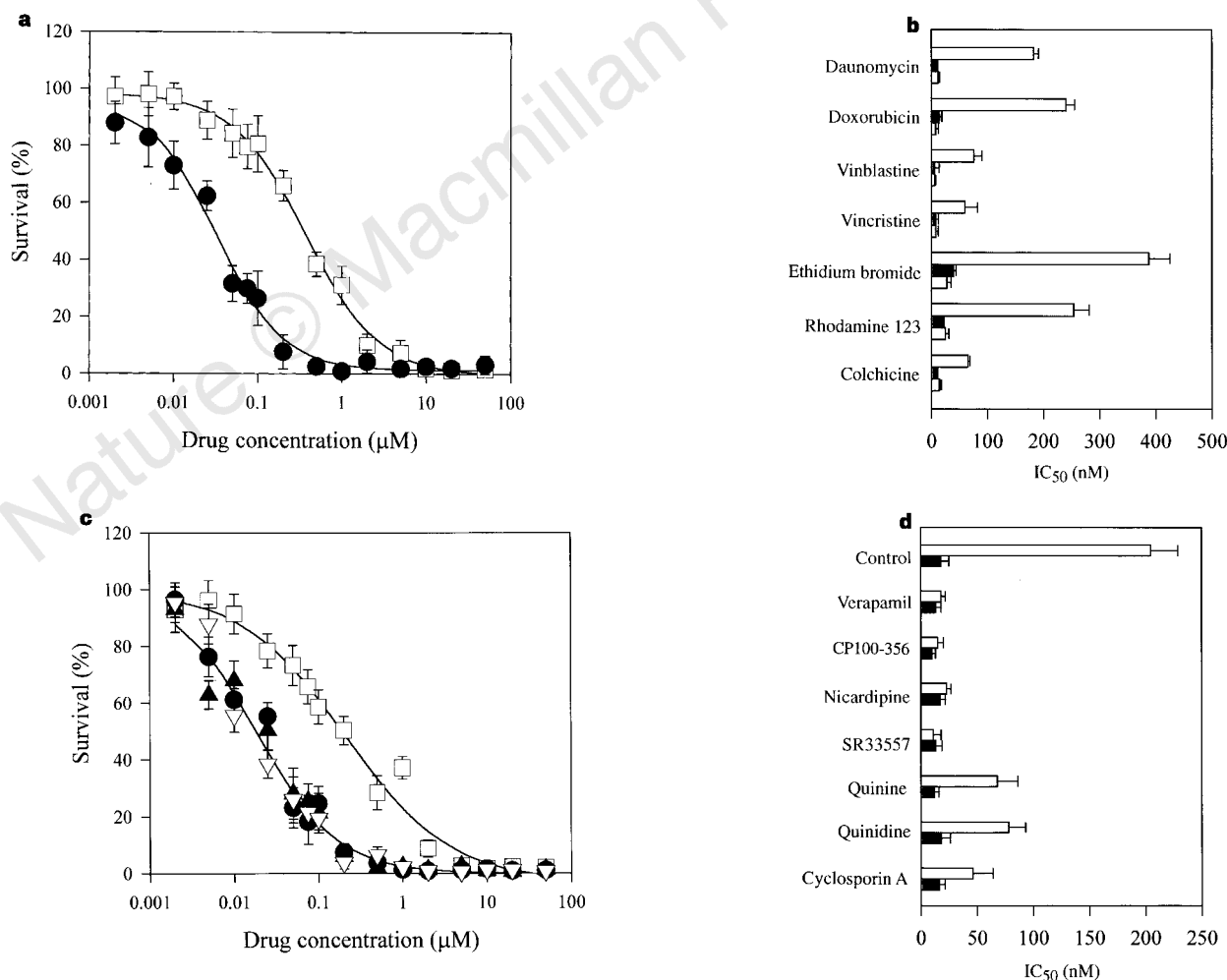


Figure 2 Drug-resistance phenotypes generated by LmrA expression in GM0637 human lung fibroblast cells. **a**, Survival of cells expressing the wild-type (squares) and K388M (circles) forms of LmrA, at increasing concentrations of ethidium bromide. **b**, Drug-survival characteristics of mock-transfected cells (grey bars), and cells expressing wild-type (white bars) and K388M (black bars) forms of LmrA. IC₅₀,

refers to the half-maximal inhibitory concentration. **c**, Daunomycin resistance of cells expressing wild-type LmrA in the absence (□) and presence (▽) of the P-glycoprotein modulator CP100-356 and the K388M form of LmrA in the absence (●) and presence (▲) of CP100-356. **d**, Effect of individual modulators on the daunomycin-resistance of cells expressing the wild-type (□) and K388M (■) forms of LmrA.

activity at 0.13 ± 0.03 mM Mg-ATP) (Fig. 4a).

Remarkably, LmrA and P-glycoprotein bound vinblastine with similar affinity (LmrA: $K_d = 46 \pm 6$ nM vinblastine; P-glycoprotein: $K_d = 39 \pm 8$ nM vinblastine) (Fig. 4b). Vinblastine binding to LmrA was inhibited by P-glycoprotein modulators, including the 1,4-dihydropyridine nicardipine and the indolizine sulphone SR33557 (Fig. 4c), which reverse LmrA-mediated drug resistance in human lung fibroblast cells (Fig. 2d).

P-glycoprotein possesses 1,4-dihydropyridine- and indolizine sulphone-selective drug-binding sites which are allosterically coupled to a vinca-alkaloid-selective binding site^{14–20}. The pharmacological interactions between these modulators and LmrA were examined by drug-dissociation kinetics¹⁵ (Fig. 4d). Nicardipine caused an acceleration of vinblastine dissociation from LmrA in a similar fashion to that observed for P-glycoprotein. The verapamil analogue CP100-356 did not affect the kinetics of vinblastine dissociation from wild-type LmrA. These results show that CP100-356 is a competitive inhibitor of vinblastine binding, and can only gain access to the vinca-alkaloid-binding site of LmrA when vinblastine has dissociated. In contrast, nicardipine acts as a non-competitive inhibitor that interacts with LmrA at a binding site distinct from, but allosterically linked to, the vinca-alkaloid-binding site. Thus, not only are the substrate and modulator specificities of LmrA and P-glycoprotein very similar, but the binding sites for substrates and allosteric non-competitive inhibitors are also conserved between P-glycoprotein and LmrA.

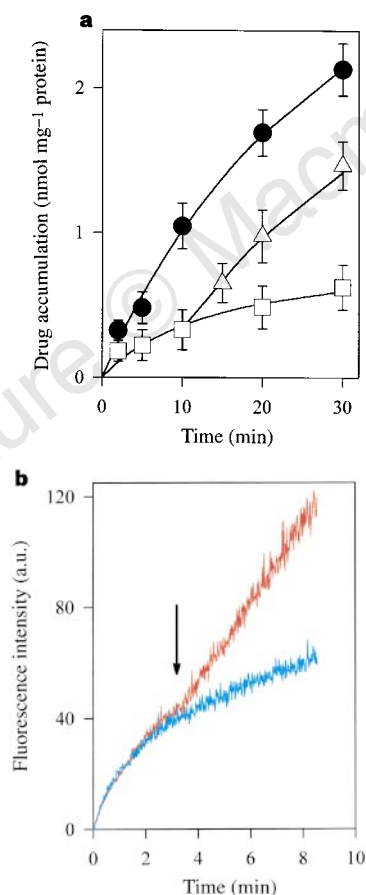


Figure 3 Wild-type LmrA protein functions as a multidrug transporter in GM0637 human lung fibroblast cells. **a**, [³H]Vinblastine accumulation by cells expressing the wild-type (□) and K388M mutant (●) forms of LmrA. After 10 min of vinblastine accumulation, verapamil was added to wild-type LmrA-expressing cells (Δ) to inhibit LmrA-mediated vinblastine extrusion. **b**, Transport of Hoechst 33342 in cells expressing wild-type LmrA (blue). Fluorescence intensity is presented in arbitrary units (a.u.). At the time indicated by the arrow, reserpine was added to one of the incubations (red) to inhibit LmrA-mediated extrusion of Hoechst 33342.

To our knowledge, the bacterial antibiotic-resistance protein LmrA is the first prokaryotic membrane protein to be functionally expressed in mammalian cells. The correct sorting of LmrA to the plasma membrane in mammalian cells is surprising in itself. Most remarkably, the drug-resistance phenotype of LmrA- and P-glycoprotein-expressing human lung fibroblast cells, and the range of reversing agents able to modulate LmrA and P-glycoprotein-dependent drug efflux were similar. Thus, the bacterial LmrA protein and the human multidrug-resistance P-glycoprotein are functionally identical. It has been proposed that P-glycoprotein participates in the protection of human cells against hydrophobic xenobiotics by active excretion of these compounds from the membrane into bile, urine or the intestinal lumen, preventing their accumulation in critical organs such as the brain²¹. Likewise, the principal physiological role of LmrA and related multidrug transporters in bacteria could involve the efflux of hydrophobic compounds that are generated intracellularly by metabolism or which are encountered in the extracellular environment. Alternatively, there may be a common endogenous substrate, such as a lipid^{22,23}, in both bacterial and mammalian cells, which has yet to be identified.

Functional, heterologous expression of LmrA in eukaryotic cells strongly implies that its ability to confer drug resistance is independent of any auxiliary proteins. LmrA is a half-molecule version of P-glycoprotein³ and, by analogy with P-glycoprotein, is likely to be active as a homodimer. LmrA transports amphiphilic drugs from the inner leaflet of the phospholipid bilayer of *L. lactis*⁴, providing strong support for the hydrophobic vacuum cleaner mode⁴ and the flippase model²⁴ for P-glycoprotein. The remarkable conservation of function, substrate specificity and the pharmacology of allosteric sites between LmrA and P-glycoprotein suggests a fundamental biological role for this type of ABC transporter in prokaryotic and eukaryotic cells. □

Methods

Genetic manipulations. The *lmrA* gene was amplified from pGKLmrA³ by PCR using the oligonucleotide 5'-CTAGTCTAGACCACCATGCATCACCATC ACCATCAGCATGACGATGACAAAGCC-3' to introduce an *Xba*I site, Kozak sequence⁹, hexa-histidine tag, and enterokinase cleavage site at the 5'-end of *lmrA*, and the oligonucleotide 5'-ATGGCCGACGTCGACTTATTATTGACC AACAGTCAATTGTTCTGAAACATATTTAGC-3' to introduce a *Sal*I site at the 3'-end of *lmrA*. The K388M mutant form of LmrA was constructed by site-directed mutagenesis of *lmrA* using the Altered Sites System (Promega) and the oligonucleotide 5'-GCCTTTGCTGGTCTCTGGTGGTGGTATGTCAAC-CATCTTCTCACTTTTAGAACG-3'. The wild-type and mutant *lmrA* genes were cloned into the mammalian expression vector pCIneo (Promega) as 1,796-base-pair (bp) *Xba*I–*Sal*I fragments, giving pCHLmrA and pCHLmrAK388M, respectively, and sequenced to ensure that only the intended changes had been introduced.

The wild-type *lmrA* gene and human *MDR1* gene¹¹ were inserted into the genome of the *Autographa californica* nuclear polyhedrosis virus under the control of the polyhedrin promoter using the BacPAK baculovirus expression system (Clontech). The wild-type *lmrA* gene was subcloned from pCIneo into the transfer vector pBacPAK9 as a 1,807-bp *Xba*I–*Not*I fragment. The human *MDR1* gene, together with a Kozak sequence⁹, was subcloned from pKSMR1 (D. Gill, unpublished data) into the transfer vector pBacPAK9 using flanking *Bam*HI and *Xho*I restriction sites. To generate recombinant baculovirus carrying the *lmrA* or *MDR1* gene, the permissive host insect cell line Sf9¹³ was cotransfected with modified transfer vector and linearized *Bsu*36 I-digested BacPAK6 viral DNA by lipofection (Clontech). After isolation of recombinant baculovirus DNA, the *lmrA* and *MDR1* genes were sequenced to confirm their proper insertion into the baculovirus genome.

Cell culture and membrane preparations. The human lung fibroblast cell line GM0637 was cultured in DMEM medium supplemented with 10% fetal calf serum¹¹. For transfection, lung fibroblast cells were seeded at a density of 2×10^5 in a 35-mm dish and grown for 18 h to ~40% confluence. Cells were washed three times with serum-free Opti-MEM (Life Technologies). 5 μg of plasmid DNA in 1 ml Opti-MEM was gently mixed with 25 μg of Lipofectin reagent (Life Technologies) in 1 ml of Opti-MEM, in a polystyrene tube. After

15 min of incubation at 20 °C, the DNA–lipid mixture was added to the washed lung fibroblast cells. Following an incubation of 5 h, the opti-MEM was replaced by DMEM supplemented with 10% fetal calf serum. Subsequently, cells were incubated for 4 h.

Transfected lung fibroblast cells were lysed in PBS containing 2% SDS. Sf9 insect cells were grown in TC100 medium (Life Technologies) supplemented with 10% fetal calf serum, and infected with wild-type or recombinant baculovirus at a multiplicity of infection of 10. Plasma membranes were prepared from baculovirus-infected insect cells, 48 h post-infection. Cells were disrupted by nitrogen cavitation at 1,500 p.s.i. The membranes were purified by sucrose density centrifugation²⁵ and stored at –70 °C in 10 mM Tris-Cl (pH 7.4) supplemented with 0.25 M sucrose.

Immunocytochemistry. Protein was analysed by SDS–PAGE (10%) followed by western blot analysis using the anti-hexa histidine tag monoclonal antibody DIA900 (Dianova). Expression of P-glycoprotein and MRP1 was assayed using the monoclonal antibody C219 and the polyclonal antibody K5, respectively. For confocal fluorescence microscopy, DIA900 antibody labelling of LmrA in cold acetone–methanol fixed fibroblast cells was visualized with anti-mouse IgG1 antibody conjugated to fluorescein (Sigma). Cells were imaged with a MRC-600 confocal microscope using a Nikon ×60 PlanoApo 1.4NA oil immersion objective lens.

Cell cytotoxicity assays. For cell cytotoxicity assays²⁶, trypsinized transfected lung fibroblast cells from various 35-mm dishes were pooled, and seeded in 24-well dishes at a density of 3×10^4 cells per well and incubated for 12 h. Cytotoxic drugs were then added to each well at the indicated concentrations.

When used, modulators (inhibitors of P-glycoprotein-mediated drug transport) were tested at 0.1 μM. After 48 h incubation, the number of viable cells in each well was estimated by methylene blue staining. Drug survival is expressed as the IC₅₀ (the concentration necessary to reduce the plating efficiency of the cells by 50% after 48 h incubation). Data represent the mean and standard error of at least two independent experiments, each carried out in duplicate.

Drug transport assays. Vinblastine accumulation by attached lung fibroblast cells in transport buffer²⁷ was initiated by addition of [³H]vinblastine (0.5×10^{12} Bq mmol^{–1} Amersham) at 100 nM. Following incubation at room temperature and washing with ice-cold PBS, cell extracts in 0.4 ml 0.4 M NaOH were neutralized with 0.8 ml 0.25 M ammonium acetate (pH 6.6). The incorporated radioactivity was estimated by liquid scintillation counting. Hoechst 33342 (Molecular Probes) accumulation by trypsinized lung fibroblast cells in 50 mM potassium HEPES (pH 7.0), 2 mM MgSO₄ and 0.9% NaCl, was measured at a drug concentration of 5 μM by fluorescence spectrometry²⁸ using excitation and emission wavelengths of 355 and 457 nm, and slit widths of 5 nm each. In drug accumulation assays, verapamil and reserpine were used at 1 μM.

ATPase assay. The ATPase activity in plasma membranes was based on a colorimetric ascorbic acid/ammonium molybdate assay²⁹ to measure the liberation of inorganic phosphate from ATP. Assays were performed in the presence of 0.1 mM EGTA and 2 mM ouabain to inhibit P-type ATPases.

Drug-binding and dissociation assays. For equilibrium binding of vinblastine¹⁵ to LmrA, plasma membranes were incubated with [³H]vinblastine at the indicated concentrations at 22 °C in the dark. After 90 min, 3 ml ice-cold

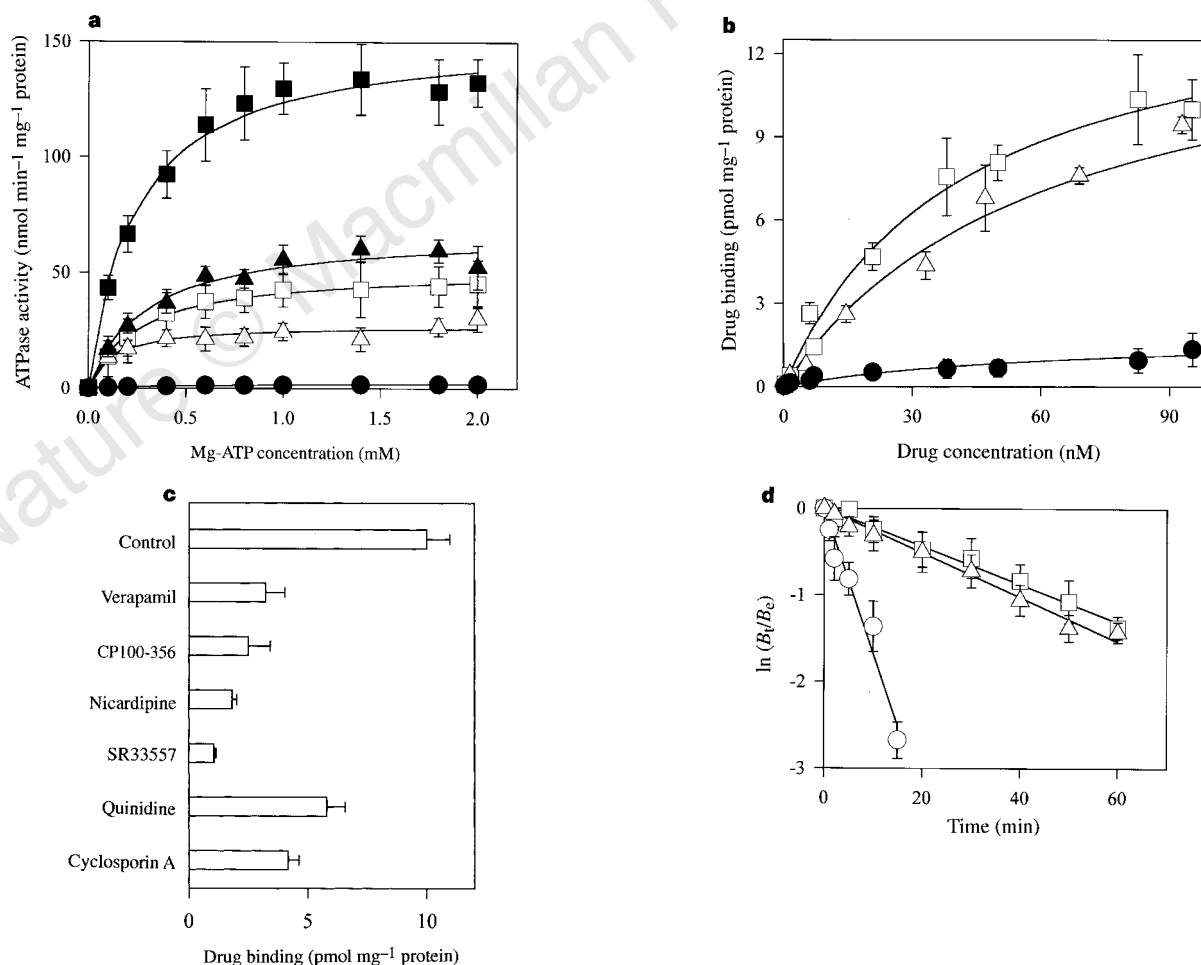


Figure 4 Pharmacological properties of wild-type LmrA and human P-glycoprotein expressed in insect cell membranes. **a**, Effect of the magnesium-ATP concentration on the basal ATPase activity of LmrA in the absence (□) or presence of 50 μM verapamil (■) or 2 mM sodium vanadate (●), and of P-glycoprotein in the absence (Δ) or presence of 50 μM verapamil (▲). **b**, Equilibrium binding of [³H]vinblastine to wild-type LmrA-containing (□), P-glycoprotein-

containing (Δ) and control (●) plasma membranes, as a function of the free concentration of [³H]vinblastine. **c**, Effect of P-glycoprotein modulators on the equilibrium binding of [³H]vinblastine to wild-type LmrA. **d**, Dissociation kinetics of [³H]vinblastine from wild-type LmrA in the absence (□) and presence of CP100–356 (Δ) or nicardipine (○).

20 mM Tris-Cl (pH 7.4) containing 2 mM MgSO_4 was added. Samples were filtered through Whatman GF/F filters. Filter-retained radioactivity was measured by liquid scintillation counting. Nonspecific binding (usually ~35% total) was defined as the amount of [^3H]vinblastine bound in the presence of 10 μM unlabelled vinblastine and was subtracted from all values. When used, modulators were included at a final concentration of 5 μM .

In drug-dissociation kinetic experiments^{15,17}, plasma membranes were equilibrated for 120 min at 22 °C in the dark, in the presence of 92 nM [^3H]vinblastine. Subsequently, the plasma membranes were taken to 12 °C after which non-labelled vinblastine was added to a final concentration of 10 μM in the presence or absence of either 3 μM nicardipine or 3 μM CP100 = 356. Samples were filtered at the indicated times as described. In Fig. 4d, dissociation of [^3H]vinblastine was plotted as the natural logarithm of the ratio of the amount of drug bound at time t (B_t) over the amount at equilibrium (B_e), as a function of time t .

Received 28 May; accepted 24 September 1997.

1. Van Veen, H. W. & Konings, W. N. Drug efflux proteins in multidrug resistant bacteria. *Biol. Chem.* **378**, 769–777 (1997).
2. Nikaido, H. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382–388 (1994).
3. Van Veen, H. W. *et al.* Multidrug resistance mediated by a bacterial homolog of the human multidrug transporter MDR1. *Proc. Natl Acad. Sci. USA* **93**, 10668–10672 (1996).
4. Bolhuis, H. *et al.* Multidrug resistance in *Lactococcus lactis*: evidence for ATP-dependent drug extrusion from the inner leaflet of the cytoplasmic membrane. *EMBO J.* **15**, 4239–4245 (1996).
5. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).
6. Chen, C. *et al.* Internal duplication and homology with bacterial transport proteins in the *mdr1* (P-glycoprotein) gene from multidrug resistance human cells. *Cell* **47**, 381–389 (1986).
7. Endicott, J. A. & Ling, V. The biochemistry of P-glycoprotein mediated multidrug resistance. *Annu. Rev. Biochem.* **58**, 137–171 (1989).
8. Gottesman, M. M., Hrycyna, C. A., Schoenlein, P. V., Germann, U. A. & Pastan, I. Genetic analysis of the multidrug transporter. *Annu. Rev. Genet.* **29**, 607–649 (1995).
9. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
10. Gill, D. R. *et al.* Separation of drug transport and chloride channel functions of the human multidrug resistance P-glycoprotein. *Cell* **71**, 23–32 (1992).
11. Goodfellow, H. R. *et al.* Protein kinase C-mediated phosphorylation does not regulate drug transport by the human multidrug resistance P-glycoprotein. *J. Biol. Chem.* **271**, 13668–13674 (1996).
12. Cole, S. P. C. *et al.* Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* **258**, 1650–1654 (1992).
13. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *In Vitro* **13**, 213–217 (1977).
14. Tamai, I. & Safa, A. R. Azidopine non-competitively interacts with vinblastine and cyclosporin A binding to P-glycoprotein in multidrug resistant cells. *J. Biol. Chem.* **266**, 16796–16801 (1991).
15. Ferry, D. R., Russell, M. A. & Cullen, M. H. P-glycoprotein possesses a 1,4-dihydropyridine-selective drug acceptor site which is allosterically coupled to a vinca-alkaloid-selective binding site. *Biochem. Biophys. Res. Commun.* **188**, 440–445 (1992).
16. Ferry, D. R. & Kerr, D. J. *In New Molecular Targets For Cancer Chemotherapy* (eds Kerr, D. J. & Workman, P.) 177–193 (CRC Press, London, 1994).
17. Malkhandi, J. *et al.* Dextrinidipine-HCl is a potent allosteric inhibitor of [^3H]vinblastine binding to P-glycoprotein of CCRF ADR 5000 cells. *Eur. J. Pharmacol.* **288**, 105–114 (1994).
18. Shao, Y.-M., Ayeshe, S. & Stein, W. D. Mutually co-operative interactions between modulators of P-glycoprotein. *Biochim. Biophys. Acta* **1360**, 30–38 (1997).
19. Martin, C., Berridge, G., Higgins, C. F. & Callaghan, R. The multidrug resistance reversal agent SR33557 modulates vinca alkaloid binding to P-glycoprotein by an allosteric interaction. *Br. J. Pharmacol.* **122**, 765–771 (1997).
20. Kostenis, E. & Mohr, K. Two-point kinetic experiments to quantify allosteric effects on radioligand dissociation. *Trends Pharmacol. Sci.* **17**, 280–283 (1996).
21. Schinkel, A. H. *et al.* Disruption of the mouse *mdr1a* P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. *Cell* **77**, 491–502 (1994).
22. Van Helvoort, A. *et al.* MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. *Cell* **87**, 507–517 (1996).
23. Ruetz, S., Brault, M., Dalton, W. & Gros, P. Functional interactions between synthetic alkyl phospholipids and the ABC transporters P-glycoprotein, Ste-6, MRP, and Pgh-1. *Biochemistry* **36**, 8180–8188 (1997).
24. Higgins, C. F. & Gottesman, M. M. Is the multidrug transporter a flippase? *Trends Biochem. Sci.* **17**, 18–21 (1992).
25. Lever, J. E. Active amino acid transport in plasma membrane vesicles from Simian virus 40-transformed mouse fibroblasts. Characteristics of electrochemical Na^+ gradient stimulated uptake. *J. Biol. Chem.* **252**, 1990–1997 (1977).
26. Cano-Cauci, D. F. & Riordan, J. R. Action of calcium antagonists on multidrug resistant cells. *Biochem. Pharmacol.* **36**, 2115–2123 (1987).
27. Callaghan, R. & Riordan, J. R. Synthetic and natural opiates interact with P-glycoprotein in multidrug resistant cells. *J. Biol. Chem.* **268**, 16059–16064 (1993).
28. Shapiro, A. B. & Ling, V. Reconstitution of drug transport by purified P-glycoprotein. *J. Biol. Chem.* **270**, 16167–16175 (1995).
29. Chifflet, S., Chiesa, U. T. R. & Tolosa, S. A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. *Analyt. Biochem.* **168**, 1–4 (1988).

Acknowledgements. We are grateful to Sahofi Recherche for supplying SR33557. We thank M. Müller for anti-MRP1 antibody K5, D. Gill for plasmid pKMDR1, and J. Taylor, K. Linton, E. Blott, G. Berridge, C. Martin, G. Begley, R. Horvath, C. Nastrocci, S. Hyde, D. Gill and P. A. McNaughton for discussions. This research was funded by the Biotechnology program of the Commission of the European Communities, the Dutch Cancer Society, the Cancer Research Campaign (UK), and the Imperial Cancer Research Fund. H.W.V.V. was the recipient of a short-term EMBO fellowship, and is a fellow of the Royal Netherlands Academy of Arts and Sciences. C.F.H. is a Howard Hughes International Research Scholar.

Correspondence and requests for materials to H.W.V.V. (e-mail: h.w.van.veen@biol.rug.nl).

The candidate tumour suppressor p33^{ING1} cooperates with p53 in cell growth control

Igor Garkavtsev*†‡, Irina A. Grigorian†‡, Valeria S. Ossovskaya‡#, Mikhail V. Chernov‡, Peter M. Chumakov‡ & Andrei V. Gudkov†

* Department of Medical Biochemistry and Southern Alberta Cancer Research Center, University of Calgary, Calgary, Alberta, Canada T2N 4N1

† Department of Molecular Genetics, College of Medicine, University of Illinois at Chicago, Chicago, Illinois 60607, USA

‡ Engelhardt Institute of Molecular Biology, Moscow, Russia

Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

¶ These authors contributed equally to this work.

The candidate tumour-suppressor gene *ING1* has been identified by using the genetic suppressor element (GSE) methodology¹. *ING1* encodes a nuclear protein, p33^{ING1}, overexpression of which inhibits growth of different cell lines. The properties of p33^{ING1} suggest its involvement in the negative regulation of cell proliferation and in the control of cellular ageing, anchorage dependence and apoptosis^{1–3}. These cellular functions depend largely on the activity of p53, a tumour-suppressor gene that determines the cellular response to various types of stress⁴. Here we report that the biological effects of *ING1* and p53 are interrelated and require the activity of both genes: neither of the two genes can, on its own, cause growth inhibition when the other one is suppressed. Furthermore, activation of transcription from the *p21/WAF1* promoter, a key mechanism of p53-mediated growth control, depends on the expression of *ING1*. A physical association between p33^{ING1} and p53 proteins has been detected by immunoprecipitation. These results indicate that p33^{ING1} is a component of the p53 signalling pathway that cooperates with p53 in the negative regulation of cell proliferation by modulating p53-dependent transcriptional activation.

Expression of *ING1* is upregulated in senescent human fibroblasts², and ectopic expression of *ING1* cDNA leads to arrest in the G1 phase of the cell cycle or induces apoptosis in several cell types^{1,3}. Inhibition of *ING1* expression by antisense RNA promotes anchorage-independent growth in mouse breast epithelial cells, increases the frequency of focus formation in NIH 3T3 cells, and prolongs the lifespan of diploid human fibroblasts in culture¹. *ING1* is mapped to human chromosome 13q34 (refs 5, 6), and loss of heterozygosity in this chromosomal locus has been reported in squamous-cell carcinomas of the head and neck⁷. Limited analysis of tumour cell lines revealed mutation of *ING1* in the neuroblastoma cell line and reduced expression in several breast carcinoma cell lines¹.

Association of *ING1* with growth suppression, replicative senescence, anchorage dependence, and apoptosis raised the question of its relationship with p53, a tumour-suppressor gene involved in the control of all these cellular functions⁴. To test this possibility, we compared the effects of overexpression or suppression of p33^{ING1} in cells differing in their p53 status. Modulation of endogenous p33^{ING1} expression was achieved by retroviral transduction of either full-length *ING1* cDNA or an antisense-oriented *ING1* cDNA fragment (anti-*ING1* GSE) acting as a potent inhibitor of p33^{ING1} expression¹. We found that the suppression of colony formation by *ING1*

Present address: ‡ Department of Functional Genomics, Genome Therapeutics Corporation, Waltham, Massachusetts 02154, USA; # Department of Pathology, University of California, San Francisco, California 94143, USA.

overexpression was evident only in the cells that maintain wild-type p53: human diploid skin fibroblasts (Fig. 1a), primary mouse embryo fibroblasts, and rat REF52 cells. Those cells that survived *ING1* transduction expressed very low levels of p33^{ING1}, as judged by immunostaining and western analysis (data not shown). No growth suppressive effect of *ING1* was detected in the cells with inactivated function of p53. Thus the *ING1*-expressing virus did not inhibit colony formation in human fibroblasts with homozygous deletion of p53 (lines MDAH041) (Fig. 1a), or in mouse embryo fibroblasts or rat REF52 cells that expressed the carboxy-terminal portion of p53 (GSE 56), a strong inhibitor of p53 function⁸ (data not shown). All of these cells were capable of expressing high levels of transduced p33^{ING1} (Fig. 2).

Another biological effect of p33^{ING1} was observed in the human fibrosarcoma cell line HT1080. Overexpression of *ING1* in these cells had only a minor negative effect on colony growth under normal conditions, but strongly increased cell sensitivity to DNA damage caused by the chemotherapeutic drug etoposide or γ irradiation. This effect was also dependent on p53, as expression of the p53-suppressor GSE56 inactivates the effect of *ING1* on drug sensitivity (Fig. 1b).

These observations indicate that the growth-inhibitory effect of *ING1* expression requires the activity of wild-type p53, and suggest that p33^{ING1} could act either upstream of or in cooperation with p53. To choose between these possibilities, we used an 'inverted' approach to analyse whether modulation of *ING1* expression could affect the growth-inhibitory function of p53. We prepared three variants of p53-deficient derivatives of Balb/c 3T3 cells (line 10(1)) that differed dramatically in p33^{ING1} content (Fig. 2a). The 10(1) cells, along with their derivatives overexpressing p33^{ING1}, are highly sensitive to ectopic expression of p53 and form rare p53-negative colonies after transduction with a retroviral vector with wild-type

p53 cDNA. However, suppression of *ING1* by anti-*ING1* GSE significantly reduced cell sensitivity to wild-type p53, as determined by the increased number of colonies formed by p53-transduced cells (Fig. 2a). Four randomly picked colonies expressed the transduced p53, unlike a few colonies that appeared in control cells (data not shown). All three variants of the 10(1) cells formed similar numbers of colonies after transduction with a control vector carrying p53^{175His}, a tumour-derived p53 mutant (Fig. 2a).

Cell rescue from the growth-inhibitory effect of p53 by *ING1* antisense RNA expression was also confirmed in co-transfection experiments. The 10(1) cells were transfected with wild-type p53 plasmid in combination with either anti-*ING1* GSE or with plain vector (Fig. 2b). The results demonstrate that p53 requires *ING1* expression to exhibit its growth-inhibitory activity, and suggest that p33^{ING1} acts in cooperation with, rather than upstream of, p53. The rescue effect of the anti-*ING1* element was much stronger than that of the anti-p53 GSE used as a positive control in this experiment.

The growth-inhibitory effect of p53 is mediated by transcriptional activation of the p53-responsive inhibitor of cyclin-dependent kinases, p21^{WAF1} (ref. 9). To analyse the effect of *ING1* expression on this important function of p53, we performed a series of experiments using a reporter expression construct in which the bacterial gene for chloramphenicol acetyltransferase (CAT) was under the control of a p53-responsive promoter with the p53-

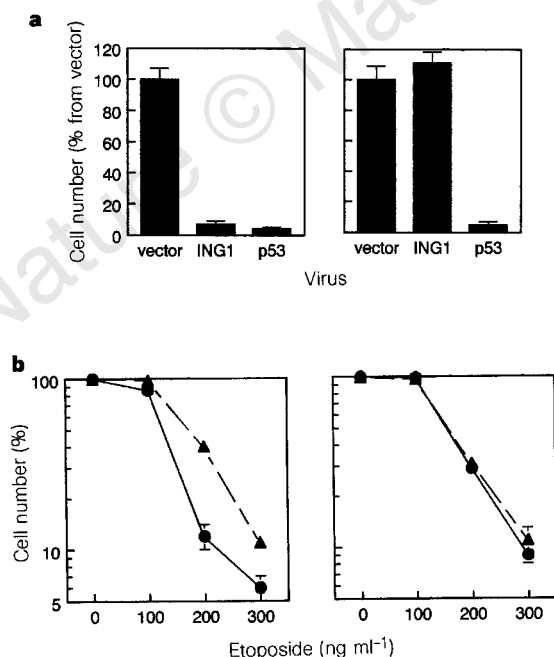


Figure 1 Biological effects of p33^{ING1} depend on p53. **a**, Two variants of human skin fibroblasts, differing in their p53 status, were infected with the indicated retroviruses, and the cell numbers were counted after selection on G418, using the MTT assay. Left, human skin fibroblasts (wild-type p53); right, MDAH041, p53-null. **b**, Wild-type HT1080 cells (left) and HT1080 with p53 suppressed by anti-p53 GSE (right) were transduced with either insert-free retroviral vector (broken lines) or with *ING1* cDNA expressing vector (solid lines), and their sensitivities to etoposide were compared by the MTT assay¹⁸.

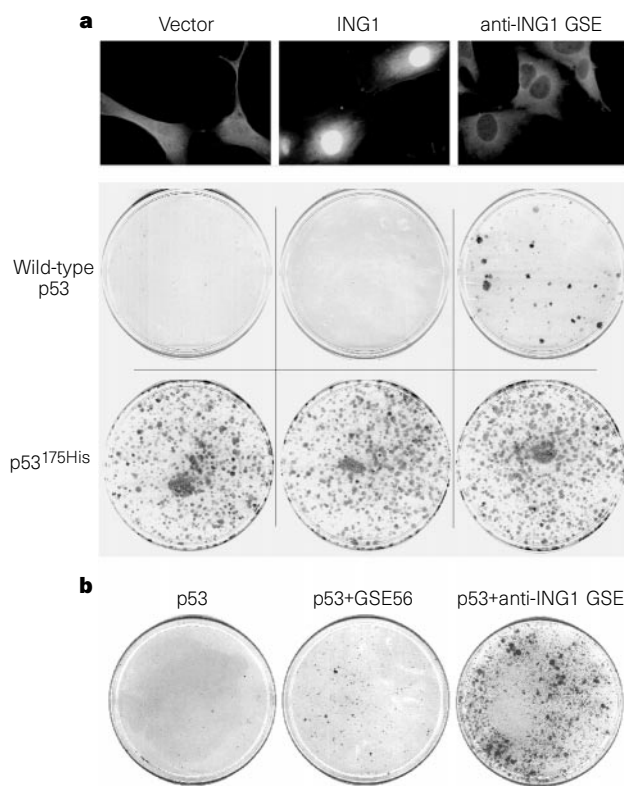


Figure 2 Growth suppression by p53 depends on p33^{ING1}. **a**, p53-null 10(1) cells, differing in p33^{ING1} expression (top row, immunofluorescent staining with anti-p33^{ING1} antibodies), were infected with retroviral vectors carrying either wild-type p53 (middle row) or the tumour-derived p53 mutant p53^{175His} (bottom row) in combination with a hygromycin-resistance gene and were selected in hygromycin. Titres of both viruses were determined by infection of p53-resistant Rat1 cells, and appeared to be approximately equal. **b**, 10(1) cells were co-transfected with the indicated combinations of plasmids (p53 refers to retroviral vector carrying wild-type p53 cDNA in combination with a hygromycin-resistance gene) and selected in hygromycin. One of three plates is shown for each transfection or infection.

binding site from the *WAF1* gene¹⁰. Two different cell lines with the wild-type *p53* gene (HT1080 and REF52) and the *p53*-deficient Balb/c 3T3 cell line 10(1) were transfected with the reporter plasmid in combination with plasmids expressing either *ING1* cDNA, anti-*ING1* GSE, a dominant-negative *p53* mutant element (GSE56), or with insert-free plasmid as a control. In the case of *p53*-deficient 10(1) cells, plasmid expressing wild-type *p53* cDNA was also added. We found that *p53*-dependent CAT activity in transfected cells was stimulated two- to fourfold in the presence of the *ING1*-expressing construct and significantly inhibited (three- to fivefold) by *ING1* antisense GSE (Fig. 3a). The inhibitory effect of the anti-*ING1* GSE was comparable to that of anti-*p53* GSE56. Similar results were obtained with the other cell lines used (data not shown). These results indicate that the function of *p53* as a transcriptional activator depends on the presence of *p33*^{ING1}, and suggest that the mechanism for the growth-inhibitory function of *p33*^{ING1} involves stimulation of *p53* transcriptional activity.

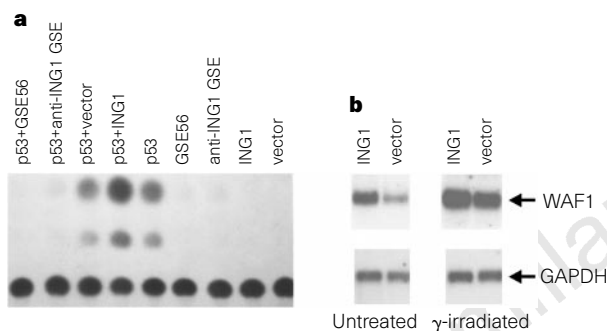


Figure 3 Expression of *p33*^{ING1} affects *p53*-dependent transactivation of *p21/WAF1*. **a**, Results of one (of three) representative CAT assay carried out with the extracts of 10(1) cells co-transfected with a combination of pWAF1-CAT plasmid and the indicated constructs. **b**, Northern blot hybridization with *p21/WAF1*-specific probe; GAPDH-specific probe was used as a loading control.

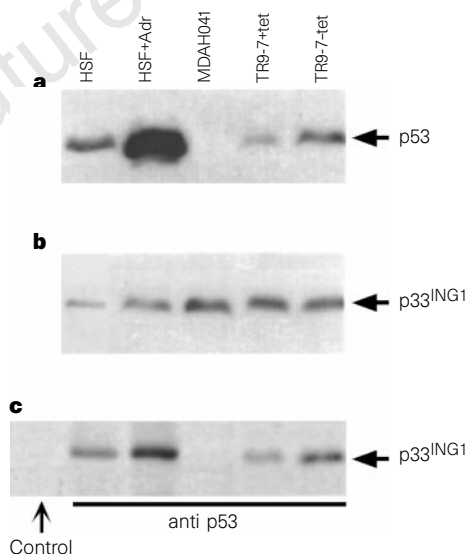


Figure 4 Co-immunoprecipitation of *p33*^{ING1} and *p53*. Detection of *p53* (**a**) and *p33*^{ING1} (**b**, **c**) proteins in the extracts (**a**, **b**) and in anti-*p53* antibody precipitates (**c**) of indicated cells by western immunoblotting. HSF, human skin fibroblasts as control. The lane contains precipitate obtained with nonspecific control antibody and HSF cell extract. +Adr, cells treated with 400 ng ml⁻¹ of Adriamycin for 6 h; MDAH041, *p53*-null cells; +tet and -tet, cells incubated with and without 1 μg ml⁻¹ tetracycline to suppress or induce *p53* expression, respectively; control.

Expression of the endogenous *p21/WAF1* gene is also affected by *ING1*: HT1080 cells overexpressing *ING1* cDNA contain 4–6 times more *p21/WAF1* mRNA than control cells do. This difference is retained even in the γ -irradiated cells that induce *WAF1* expression by a *p53*-dependent mechanism (Fig. 3b).

The possibility of a physical interaction between *p33*^{ING1} and *p53* was tested in co-immunoprecipitation experiments. Anti-*p53* antibodies were used to precipitate *p53*-containing protein complexes from cellular extracts. The presence of *p33*^{ING1} protein in precipitates was monitored by immunoblotting with polyclonal antibodies against *p33*^{ING1} (Fig. 4). We used several cell types differing in *p53* expression, including: wild-type human skin fibroblasts (HSF), growing under normal conditions or treated with the chemotherapeutic drug Adriamycin to induce *p53* stabilization (Fig. 4a); human fibroblasts from a patient with Li-Fraumeni syndrome (line MDAH041) that lack *p53*; and the derivatives of MDAH041 carrying tetracycline-regulated wild-type *p53* cDNA (TR9-7). All these cells expressed similar levels of *p33*^{ING1} (Fig. 4a, b), indicating that *p33*^{ING1} expression was not affected by DNA damage or by *p53* expression. *p33*^{ING1} protein was detected in anti-*p53* antibody precipitates in all the cells expressing wild-type *p53*, and the amount of *p33*^{ING1} in precipitates correlated with the *p53* content. In contrast, no *p33*^{ING1} was found in anti-*p53* antibody precipitates from the extracts of *p53*-null MDAH041 cells (Fig. 4c). Polyclonal antibodies against *p33*^{ING1} co-precipitated a protein of relative molecular mass 53,000 (*M*_r 53K) from the extract of ³⁵S-methionine-labelled cells expressing wild-type *p53*, but not from the *p53*-null cell (data not shown), confirming that there was a physical association between *p53* and *p33*^{ING1}.

Taken together, our results demonstrate that *p33*^{ING1} directly cooperates with *p53* in growth regulation by modulating the ability of *p53* to act as a transcriptional activator. Reduction of *ING1* expression inhibits the growth-suppressive activity of *p53*, suggesting that *p33*^{ING1} is essential for *p53* function. The mechanism of the cooperation between *p33*^{ING1} and *p53* involves physical interaction between these two proteins, which form a joint complex detectable by immunoprecipitation. These data place *p33*^{ING1} into a family of *p53*-interacting proteins, including Mdm2, Ref-1 and p300, which modulate *p53* activity through physical interaction^{11–13}. Whether *p33*^{ING1} is directly involved in the regulation of *p53* DNA binding or whether their interaction occurs at some other level (such as *p53* nuclear transport) is presently unknown and is the subject of a separate study. The involvement of *p33*^{ING1} in the *p53* signalling pathway indicates that *ING1* is a potential tumour-suppressor gene, the loss or inactivation of which may contribute to altered cell growth control, resistance to apoptosis, or establishment of the immortal phenotype in tumours retaining wild-type *p53*. □

Methods

Plasmids, antibodies and cells. Retroviral vectors expressing *ING1* cDNA, anti-*ING1* GSE, wild-type *p53* cDNA, *p53*^{175His} cDNA, and anti-*p53* GSE56 in combination with either *neo* or *hygro* selectable markers have been described^{1,8,14}. In all these vectors, except anti-*ING1* GSE, cDNA inserts were expressed from a Mo-MuLV LTR promoter and selectable marker genes were expressed from internal SV40 promoter. Anti-*ING1* GSE was expressed from a cytomegalovirus promoter. The structure of a reporter pWAF1-CAT plasmid has been described¹⁰. It contains the *CAT* gene under the control of the combination of *p53*-binding site from the *WAF1* promoter with the minimal heat-shock Hsp70 promoter. Hybridomas producing monoclonal antibodies against *p53*, PAb421 and DO1, as well as *p53*-negative Balb/c 3T3 cells [line 10(1)], were provided by A. Levine¹⁶. The human fibrosarcoma cell line HT1080 expressing ecotropic retroviral receptor, as well as *p53*-negative Li-Fraumeni fibroblasts (line MDAH01) and their derivative expressing wild-type *p53* under the control of a tetracycline-regulated promoter (line TR9-7)¹⁷ were provided by G. Stark. All cell lines were maintained in DMEM with 10% fetal

calf serum. TR9-78 cells were maintained continuously in the presence of $1 \mu\text{g ml}^{-1}$ tetracycline to suppress expression of p53; to induce p53 expression, cells were incubated without tetracycline for 24 h.

Transfection and retroviral infection. Delivery of the cDNAs and GSEs into different cell types was performed by calcium phosphate-mediated transfection or retroviral transduction as described⁸. The infected cells were selected on G418 ($500 \mu\text{g ml}^{-1}$) or hygromycin ($200 \mu\text{g ml}^{-1}$) for 7–14 days. All the experiments were reproduced at least 3 times using 2–4 parallel plates for each DNA or virus.

Western analysis. Western analysis was performed using rabbit polyclonal antibodies against a bacterially expressed glutathione S-transferase (GST)–p33^{ING1} fusion protein¹ or a mixture of anti-p53 monoclonal antibodies PAB421 and DO1 (provided by A. Levine). Biotinylated donkey anti-rabbit or sheep anti-mouse antibodies were used as secondary antibodies. Antibody binding was visualized by enhanced chemiluminescence using horseradish peroxidase conjugated with streptavidin (Amersham).

Immunoprecipitation analysis. Cells growing in 100-mm tissue-culture dishes (approximately 3×10^6 cells per dish) were washed with ice-cold PBS, scraped into 1 ml of RIPA buffer with no SDS and sonicated. Extracts were cleared by centrifugation at $10,000g$ for 10 min. A mixture of PAB421 and DO-1 monoclonal antibodies was added to cell extract and incubated for 2 h at 4°C . Protein A Sepharose ($30 \mu\text{l}$) equilibrated in RIPA was then added and incubated for a further 30 min. The beads were extensively washed with ice-cold RIPA and the precipitate was dissolved in a sample buffer for electrophoresis and western analysis.

The p53 transactivation assay. This was carried out as described¹⁰, with 2×10^5 cells per 60-mm dish being transfected with $12 \mu\text{g}$ of plasmid DNA containing $4 \mu\text{g}$ of pWAF1-CAT and $2 \mu\text{g}$ of pCMV-lacZ. Cell extracts prepared by freezing and thawing were normalized for protein content. The efficiency of transfection was determined using a quantitative β -galactosidase assay.

Drug sensitivity assay. HT1080 cells (10^5 cells per well of a 6-well plate) transduced with different retroviral constructs were incubated in the presence of different concentrations of etoposide for 4 days. Cell viability was then determined using the MTT assay¹⁸. The experiment was repeated three times using three parallel wells for each drug concentration.

Received 15 July; accepted 17 October 1997.

- Garkavtsev, I. A., Kazarov, A. R., Gudkov, A. V. & Riabowol, K. Suppression of the novel growth inhibitor p33^{ING1} promotes neoplastic transformation. *Nature Genet.* **14**, 415–420 (1996).
- Garkavtsev, I. & Riabowol, K. Extension of the replicative life span of human diploid fibroblasts by inhibition of the p33^{ING1} candidate tumor suppressor. *Mol. Cell Biol.* **17**, 2014–2019 (1997).
- Helbing, C. C., Veillette, C., Riabowol, K., Johnston, R. N. & Garkavtsev, I. A novel candidate tumor suppressor, *ING1*, is involved in the regulation of apoptosis. *Cancer Res.* **57**, 1255–1258 (1997).
- Gottlieb, T. M. & Oren, M. p53 in growth control and neoplasia. *Biochim. Biophys. Acta* **1287**, 77–102 (1996).
- Garkavtsev, I., Demetrick, D. & Riabowol, K. Cellular localization and chromosome mapping of a novel candidate tumor suppressor gene. *Cytogenet. Cell Genet.* **76**, 176–178 (1997).
- Zeremski, M., Horrigan, S., Grigorian, I. A., Westbrook, C. & Gudkov, A. V. Chromosomal localization of p33^{ING1} candidate tumor suppressor gene. *Somat. Cell Mol. Genet.* **23**, 233–236 (1997).
- Maestro, R. et al. Chromosome 13q deletion mapping in head and neck squamous cell carcinomas: identification of two distinct regions of preferential loss. *Cancer Res.* **56**, 1146–1150 (1996).
- Ossovskaya, V. S. et al. Dissection of p53 functions by genetic suppressor elements: distinct biological effects of separate p53 domains. *Proc. Natl Acad. Sci. USA* **93**, 10309–10314 (1996).
- El-Deiry, W. S. et al. WAF-1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Kondratov, R. V. et al. Functional heterogeneity of p53-responsive elements. *Mol. Biol. (Mosk.)* **30**, 613–620 (1996).
- Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The *mdm-2* oncogene product forms a complex with the p53 protein and inhibits p53 mediated transactivation. *Cell* **69**, 1237–1245 (1992).
- Jayaraman, L. et al. Identification of redox/repair protein Ref-1 as a potent activator of p53. *Genes Dev.* **11**, 558–570 (1997).
- Avantaggiati, M. L. et al. Recruitment of p300/CBP in p53-dependent signal pathway. *Cell* **89**, 1175–1184 (1997).
- Kopin, B. P. et al. Influence of exogenous ras and p53 on P-glycoprotein function in immortalized rodent fibroblasts. *Oncol. Res.* **7**, 299–306 (1995).
- Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
- Harvey, D. M. & Levine, A. J. p53 alteration is a common event in the spontaneous immortalization of primary BALB/c murine embryo fibroblasts. *Genes Dev.* **5**, 2375–2385 (1991).
- Agarwal, M. L., Agarwal, A., Taylor, W. R. & Stark, G. R. p53 controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts. *Proc. Natl Acad. Sci. USA* **92**, 8493–8497 (1995).
- Pauwels, R. et al. Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds. *J. Virol. Methods* **20**, 309–321 (1988).

Acknowledgements. We thank R. Davidson and E. Feinstein for reading the manuscript, and C. Zelnick for help with manuscript preparation. This work was supported by NIH grants to A.V.G., by a Fogarty International Collaboration Award to A.V.G. and P.M.C., and fellowships from NCI to I.A.G. and V.S.O.

Correspondence and requests for materials should be addressed to A.V.G. (e-mail: gudkov@uic.edu).

MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes

Timothy J. Hemesath, E. Roydon Price, Clifford Takemoto, Tina Badalian & David E. Fisher

Division of Pediatric Hematology/Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

Germline mutations at loci encoding the transcription factor Microphthalmia (Mi), the cytokine receptor c-Kit, or its ligand Steel factor (Sl) result in strikingly similar defects in mast cell and melanocyte development^{1–3}. Here we describe a biochemical link between Kit signalling and the activity of Mi. Stimulation of melanoma cells with Sl results in activation of MAP kinase, which in turn phosphorylates Mi at a consensus target serine. This phosphorylation upregulates Mi transactivation of the tyrosinase pigmentation gene promoter. In addition to modulating pigment production, such signalling may regulate the expression of genes essential for melanocyte survival and development. The pathway represents a new application of the general MAP kinase machinery in transducing a signal between a tissue-specific receptor at the cell surface and a tissue-specific transcription factor in the nucleus.

Microphthalmia (Mi) is a transcription factor of the basic, helix–loop–helix, leucine-zipper (bHLH) type, the mutation of which causes Waardenburg Syndrome type II in humans⁴. In mice, a profound loss of pigmented cells in the skin, eye and inner ear results, as well as osteopetrosis and defects in natural killer and mast cells^{5,6}. Mi binds E-box-type enhancer elements and may heterodimerize with the related family members TFEB, TFEC and TFE3 (ref. 7). Mutations in c-Kit or its ligand Sl (stem-cell factor, mast-cell growth factor) likewise result in animals lacking melanocytes and functional mast cells, together with defects in haematopoiesis and germ-cell development^{2,3}. The striking phenotypic overlap has led to suggestions that Sl, c-Kit and Mi function in a common growth or differentiation pathway^{8,9}.

Western blot analysis of a human melanoma cell line revealed two Mi species with relative mobilities corresponding to relative molecular masses of 54K and 60K. Activation of c-Kit by Sl completely shifted the lower band to the position of the upper band (Fig. 1a). This shift occurred rapidly but transiently; the lower band reappeared within two hours. A similar but sustained shift was induced when cells were treated with the phorbol ester TPA (12-O-tetradecanoylphorbol-13-acetate), a potent activator of protein kinase C. Using phosphotyrosine antibodies, we found that treatment with Sl but not TPA led to the expected phosphorylation of c-Kit (Fig. 1a). Both stimuli resulted in the activation of MAP kinase (MAPK), which correlated temporally with the shift in Mi protein (Fig. 1a). *In vitro* phosphatase treatment resulted in a discrete, dose-dependent shift of the upper form of Mi to the faster-migrating species (Fig. 1b). This shift was blocked by phosphatase inhibitors, suggesting that the mobility change observed in cell extracts was due to phosphorylation.

In vivo labelling of the Mi proteins with [³²P]-orthophosphate indicated that both Mi forms were phosphoproteins (see below). Phosphoamino-acid analysis revealed phosphoserine, but no detectable phosphotyrosine or phosphothreonine (Fig. 2a), a finding consistent with the failure of anti-phosphotyrosine antibodies to detect Mi in total cell extracts (not shown). These results indicated that c-Kit was not responsible for phosphorylating Mi directly, so we focused on downstream kinases.

Mi. It is likely that the specificity of the MAPK pathway between Kit and Mi lies in the highly restricted temporal and spatial expression of Sl, Kit and Mi *in vivo*. Mi/Kit signalling illustrates application of general signalling machinery to link an individual cytokine to a specific transcription factor (Fig. 4b). These results are consistent with a report that subcutaneous Sl induces localized pigmentation in humans¹⁸. An increase in cyclic AMP may also upregulate the tyrosinase promoter in a manner involving Mi by means of signals originating from the receptor for melanocyte-stimulating hormone^{19,20}.

Mast cells from *mi* mutant mice underexpress c-Kit²¹, and Mi may upregulate c-Kit expression from a binding site in the c-Kit promoter²². Also, transfection of the c-Kit-related CSF-1 receptor restores cytokine responsiveness to *kit*-mutant but not *mi*-mutant mast cells⁹, consistent with the role proposed here for Mi as a downstream target of cytokine-initiated MAPK signals. Moreover, genetic experiments^{23,24} demonstrated that germline deficiency of the phosphatase SHP1 (a negative regulator of Kit signalling) enhances Sl-induced MAPK activation and partially rescues the number of resident mast cells. Kit-stimulated MAPK activation and

Mi phosphorylation may alter expression of genes controlling cell lineage commitment, development or survival. □

Methods

Immunoprecipitation and western blot analysis. The monoclonal antibody D5 was raised against a histidine fusion protein expressed from the amino-terminal *Taq-Sac* fragment of human Mi (MITF) cDNA²⁵ and produces a specific gel-mobility supershift with Mi, but not with the related proteins TFEB, TFEC and TFE3 (not shown). 501mel cells (gift from R. Halaban) were maintained in F10 medium plus 10% fetal calf serum (FCS). Cells were stimulated with 20 ng ml⁻¹ recombinant human Sl (R & D Systems) or 10 ng ml⁻¹ TPA for 8 min at 37°C unless indicated. Cells were lysed in 50 mM Tris, pH 7.6, 150 mM NaCl, 10% Triton-X 100 plus protease and phosphatase inhibitors. Samples were solubilized in SDS-sample buffer plus 0.5% 2-mercaptoethanol and boiled for 5 min. Following SDS-PAGE and transfer to nitrocellulose, blots were blocked in 5% milk/0.05% Tween-20 in Tris-buffered saline before antibody incubation. Proteins were detected with peroxidase-conjugated second-step antibody (Cappel) and chemiluminescence reagents (Amersham).

Phosphatase treatment. Immunoprecipitated Mi was washed three times with lysis buffer, twice with buffer A (100 mM KCl, 20 mM HEPES, pH 7.4, 0.2 mM EDTA, 2 mM DTT, plus protease inhibitors), and resuspended in 40 µl buffer A. Control digests contained phosphatase inhibitors sodium vanadate (1 mM), sodium pyrophosphate (20 mM), and sodium fluoride (10 mM). Potato acid phosphatase (Boehringer Mannheim) was added to immunoprecipitates (IPs) for 15 min at 30°C. The reaction was stopped with phosphatase inhibitors and analysed by western blot.

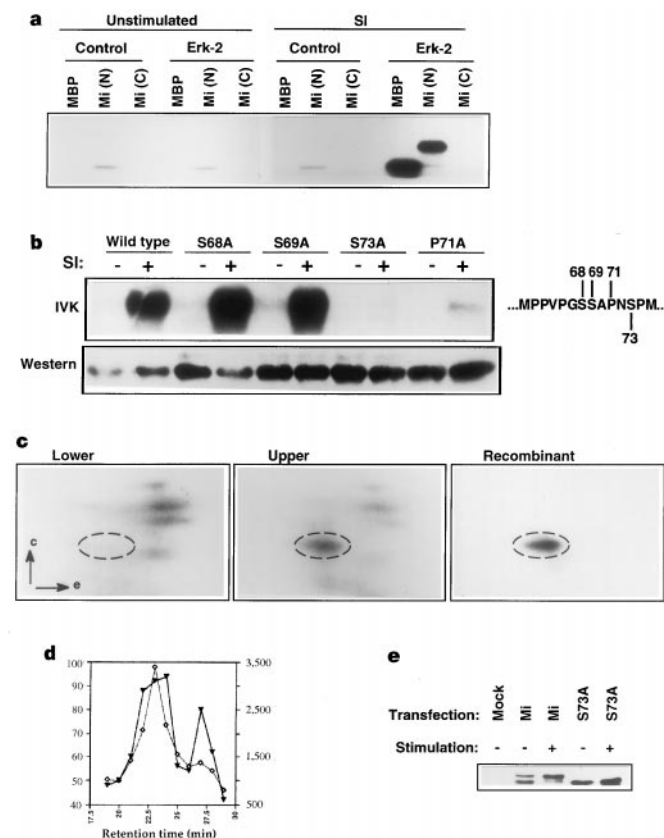


Figure 3 Mi phosphorylation at S73 by MAPK. **a**, Lysates from $\pm$ SI-stimulated melanoma cells were immunoprecipitated (anti-Erk-2 or control) and tested in IVK using N-terminal (N) or C-terminal (C) recombinant Mi substrates or myelin basic protein (MBP). **b**, Mutations in Mi were tested as IVK substrates from $\pm$ SI cell extracts immunoprecipitated with anti-ERK and identified Ser73 as phosphoacceptor. **c**, Two-dimensional tryptic maps from ³²P-labelled endogenous cellular Mi lower and upper bands revealed a distinct Kit-dependent spot (hatched circle) which co-migrates with ERK/IVK phosphorylated Mi. 'c' and 'e', Chromatographic and electrophoretic migration, respectively. **d**, HPLC fractionation of *in vivo* ³²P-labelled Mi tryptic digests from the upper band (diamonds) shows co-elution with ERK/IVK recombinant Mi (triangles); y-axes show beta counts. The secondary peak probably results from oxidative peptide-bond cleavage from performic acid¹⁴. **e**, Wild-type or S73A mutant Mi were transfected into COS-7 cells, followed by TPA stimulation as indicated. S73A fails to undergo mobility shift.

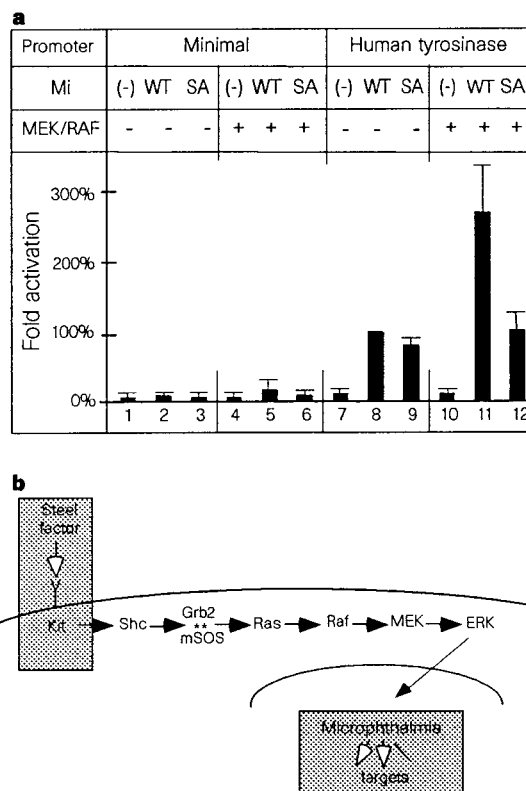


Figure 4 MAPK phosphorylation enhances Mi-dependent transactivation. **a**, Wild-type or S73A-mutant Mi were co-transfected with minimal or tyrosinase promoter-driven luciferase reporters into BHK cells. Constitutively active Raf and wild-type MEK were included as indicated to activate the MAPK pathway. Mean fold activation (normalized to 100% for activity of wild-type Mi in the absence of Raf/MEK, column 8) from three independent experiments is shown with standard error bars. Raf/MEK potentiates transactivation by wild-type but not S73A Mi. **b**, Model for signal transduction from Kit to Mi. Tissue-specific factors are shown in grey boxes. Asterisks indicate non-covalent association between the adaptor protein Grb2 and the nucleotide exchange factor mSOS. Shc is an adaptor protein that links c-Kit to the Ras/MAPK pathway.

Phosphoamino-acid analysis, tryptic mapping, HPLC fractionation. 501mel cells were starved for 30 min in serum-free, phosphate-free RPMI medium, then labelled for 3 h using 1 mCi ml^{-1} ^{32}P -labelled inorganic phosphate. Cells were stimulated with SI or TPA and solubilized in IP buffer. Mi proteins were immunoprecipitated overnight at 4°C , electrophoresed and transferred to nitrocellulose. Bands were cut out and digested for 20 h at 37°C with $25 \mu\text{g}$ TPKC-treated trypsin (Sigma). Phosphoamino-acid analysis and phosphopeptide mapping were carried out as described¹⁴ using ammonium carbonate at pH 8.9. For HPLC fractionation, a 25-cm C18 reverse-phase column (Vydac) and an acetonitrile gradient (0–70% in 0.1% trifluoroacetic acid) were used at a flow rate of 0.2 ml min^{-1} . Fractions were assayed by Cerenkov counting.

In vitro kinase assay. Cells were activated and lysed and $4 \mu\text{l}$ of anti-Erk-2 antiserum (Santa Cruz) plus $20 \mu\text{l}$ protein-A-agarose beads were added to the lysate and mixed at 4°C overnight. Beads were washed three times with lysis buffer and once with IVK buffer (50 mM HEPES, pH 7.6, 2 mM sodium vanadate, 10 mM magnesium chloride, 1 mM PMSF, 2 mM DTT and $50 \mu\text{M}$ ATP). For each reaction, $40 \mu\text{l}$ IVK buffer, $1 \mu\text{l}$ $[\gamma\text{-}^{32}\text{P}]\text{ATP}$, and phospho-acceptor protein were added. Myelin basic protein ($5 \mu\text{g}$) or Mi histidine fusion proteins spanning residues 16–185 or 139–419 (ref. 25) ($4 \mu\text{l}$ of a solution having an absorbance of 0.065 at 280 nm) were added as substrates and incubated at 30°C for 30 min. Reactions were stopped by addition of $2\times$ SDS-sample buffer and analysed by western blot and autoradiography.

Luciferase assay. The human tyrosinase promoter reporter encompasses nucleotides –300 to +80 (ref. 16) in the pGL2Basic luciferase reporter (Promega). Wild-type Mi and the S73A mutant were cloned into the pEF-BOS expression vector²⁶. The plasmid encoding constitutively active Raf was a 24G deletion mutant²⁷ (gift from G. Cooper). Wild-type MEK plasmid was a gift from L. Zon. Transfections were done by adding plasmid DNA ($10 \mu\text{g}$ per 6-cm plate) to $300 \mu\text{l}$ DMEM, mixing 1:1 with a 5% lipofectamine/DMEM solution, and incubating at room temperature for 1 h. BHK cells maintained in DMEM/10% FCS were washed twice with serum-free DMEM before transfection. DNA/lipofectamine was added to 2 ml DMEM on a 6-cm plate. Cells were incubated overnight at 37°C and fed the next morning, then collected for assay 8 h later and analysed with a Monolight 2010 luminometer using reagents as recommended by the manufacturer (Analytical Luminescence). Luciferase data were normalized to β -galactosidase activity in all samples.

Received 5 June; accepted 23 September 1997.

- Moore, K. J. Insight into the microphthalmia gene. *Trends Genet.* **11**, 442–448 (1995).
- Russell, E. Hereditary anemias of the mouse: a review for geneticists. *Adv. Genet.* **20**, 357–459 (1979).
- Witte, O. Steel locus defines new multipotent growth factor. *Cell* **63**, 5–6 (1990).
- Tassabehji, M., Newton, V. E. & Read, A. P. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nature Genet.* **8**, 251–255 (1994).
- Hodgkinson, C. A. *et al.* Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395–404 (1993).
- Steingrimsson, E. *et al.* Molecular basis of mouse microphthalmia (mi) mutations helps explain their developmental and phenotypic consequences. *Nature Genet.* **8**, 256–263 (1994).
- Hemesath, T. J. *et al.* microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. *Genes Dev.* **8**, 2770–2780 (1994).
- Isozaki, K. *et al.* Cell type-specific deficiency of c-kit gene expression in mutant mice of mi/mi genotype. *Am. J. Path.* **145**, 827–836 (1994).
- Dubreuil, P. *et al.* The c-fms gene complements the mitogenic defect in mast cells derived from mutant W mice but not mi (microphthalmia) mice. *Proc. Natl Acad. Sci. USA* **88**, 2341–2345 (1991).
- Okuda, K. *et al.* Granulocyte-macrophage colony-stimulating factor, interleukin-3, and steel factor induce rapid tyrosine phosphorylation of p42 and p44 MAP kinase. *Blood* **79**, 2880–2887 (1992).
- Marshall, C. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
- Treisman, R. Regulation of transcription by MAP kinase cascades. *Curr. Opin. Cell Biol.* **8**, 205–215 (1996).
- Ahn, N., Seger, R. & Krebs, E. The mitogen-activated protein kinase activator. *Curr. Opin. Cell Biol.* **4**, 992–999 (1992).
- Boyle, W., van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **210**, 110–149 (1991).
- Hearing, V. & Jiminiez, M. Mammalian tyrosinase—the critical regulatory control point in melanocyte pigmentation. *J. Biochem.* **19**, 1141–1147 (1987).
- Bentley, N. J., Eisen, T. & Goding, C. R. Melanocyte-specific expression of the human tyrosinase promoter: activation by the microphthalmia gene product and role of the initiator. *Mol. Cell. Biol.* **14**, 7996–8006 (1994).
- Yasumoto, K., Yokoyama, K., Shibata, K., Tomita, Y. & Shibahara, S. Microphthalmia-associated transcription factor as a regulator for melanocyte-specific transcription of the human tyrosinase gene [erratum]. *Mol. Cell. Biol.* **15**, 1833 (1995). *Mol. Cell. Biol.* **14**, 8058–8070 (1994).
- Costa, J. *et al.* Recombinant human stem cell factor (kit ligand) promotes human mast cell and melanocyte hyperplasia and functional activation in vivo. *J. Exp. Med.* **183**, 2681–2686 (1996).
- Englaro, W. *et al.* Mitogen-activated protein kinase pathway and AP-1 are activated during cAMP-induced melanogenesis in B-16 melanoma cells. *J. Biol. Chem.* **270**, 24315–24320 (1995).
- Bertolotto, C., Bille, K., Ortonne, J. & Ballotti, R. Regulation of tyrosinase gene expression by cAMP in B16 melanoma cells involves two CATGTG motifs surrounding the TATA box: implication of the microphthalmia gene product. *J. Cell Biol.* **134**, 747–755 (1996).
- Ebi, Y. *et al.* Low c-kit expression of cultured mast cells of mi/mi genotype may be involved in their

- defective responses to fibroblasts that express the ligand for c-kit. *Blood* **80**, 1454–1462 (1992).
- Tsujiimura, T. *et al.* Involvement of transcription factor encoded by the mi locus in the expression of c-kit receptor tyrosine kinase in cultured mast cells of mice. *Blood* **88**, 1225–1233 (1996).
- Paulson, R., Vesey, S., Siminovich, K. & Berstein, A. Signalling by the W/Kit receptor tyrosine kinase is negatively regulated by the protein tyrosine phosphatase Shp1. *Nature Genet.* **13**, 309–315 (1996).
- Lorenz, U. *et al.* Genetic analysis reveals cell type-specific regulation of receptor tyrosine kinase c-Kit by the protein tyrosine phosphatase SHP1. *J. Exp. Med.* **184**, 1111–1126 (1997).
- Tachibana, M. *et al.* Cloning of MITF, the human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1–p12.3. *Hum. Mol. Genet.* **3**, 553–557 (1994).
- Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322 (1990).
- Stanton, V., Nichols, D., Laudano, A. & Cooper, G. Definition of the human raf amino-terminal regulatory region by deletion mutagenesis. *Mol. Cell. Biol.* **9**, 639–647 (1989).

Acknowledgements. We thank J. Jackson and J. Abraham for technical help, S. Galli, G. Cooper, M. Greenberg, B. Neel, D. Ron and P. Sharp for discussion, R. Halaban for 501mel cells, and members of the Burakoff laboratory for advice and assistance. This work was supported by grants from the NIH, the Pew Foundation, and the James S. McDonnell Foundation. T.J.H. is a Medical Foundation Fellow; D.E.F. is Nirenberg Fellow.

Correspondence and requests for materials should be addressed to D.E.F. (e-mail: david_fisher@dfci.harvard.edu).

Engineering cyclophilin into a proline-specific endopeptidase

Eric Quéménéur, Mireille Moutiez, Jean-Baptiste Charbonnier & André Ménez

Département d'Ingénierie et d'Etudes des Protéines, CEA Saclay, 91191 Gif-sur-Yvette, France

Designing an enzyme requires, among a number of parameters, the appropriate positioning of catalytic machinery within a substrate-binding cleft. Using the structures of cyclophilin–peptide complexes^{1–4}, we have engineered a new catalytic activity into an *Escherichia coli* cyclophilin by mutating three amino acids, close to the peptide binding cleft, to form a catalytic triad similar to that found in serine proteases. In conjunction with cyclophilin's specificity for proline-bearing peptides, this creates a unique endopeptidase, cyproase 1, which cleaves peptides on the amino-side of proline residues. When acting on an Ala-Pro dipeptide, cyproase 1 has an efficiency (k_{cat}/K_m) of $0.7 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ and enhances the rate of reaction ($k_{\text{cat}}/k_{\text{uncat}}$) 8×10^8 -fold. This activity depends upon a deprotonated histidine and is inhibited by nucleophile-specific reagents, as occurs in natural serine proteases. Cyproase 1 can hydrolyse a protein substrate with a proline-specific endopeptidase activity.

It is anticipated that protein engineering will allow the design of enzymes with new activities adapted to social and economical needs. However, only a few successes have paved this difficult avenue^{5,6}, probably because of the need for multiple complex determinants to be adjusted adequately for an efficient transition-state stabilization to occur⁷. These determinants include a substrate-binding cleft, the catalytic machinery and various parameters such as internal molecular dynamics and water molecules to mediate enzyme–ligand interactions. With the intention of engineering a proline-specific protease we tentatively grafted a proteolytic machinery into cyclophilin, which already possesses a binding cleft for peptidyl-prolyl substrates. Cyclophilins catalyse *cis*–*trans* isomerization of X-Pro bonds^{8,9} but have never been reported to hydrolyse X-Pro bonds. As suggested by subtilisin mutants¹⁰, the presence of a single nucleophilic serine at an appropriate distance from the carbonyl carbon of the substrate amide bond may produce some proteolysis, although presumably with a weak k_{cat} ($\leq 10^{-3} \text{ s}^{-1}$). Therefore, knowing the organization of X-Pro peptide-binding clefts of free¹ and peptide-bound cyclophilins^{2–4}, we mutated into serine a number of individual residues of the X-Pro binding pocket of ECypP, a cyclophilin from *E. coli*, and searched for a possible peptidase activity.

In the crystal structures^{2–4}, the β -carbons of R48, Q56, A91 and

T93 (as numbered in ECypP), are separated from the peptidyl X-Pro carbonyl carbon by distances ranging from 4.3 Å to 8.5 Å (Fig. 1). Although larger than the corresponding distance (3.2 Å) in serine protease-inhibitor complexes¹¹, we thought that these distances might accommodate any structural change that may result from mutations of selected residues into serine and may be compatible with any of the possible *cis* and/or *trans*-proline conformations of bound substrates. In this respect, several situations have been previously described which include the presence of a *cis* conforma-

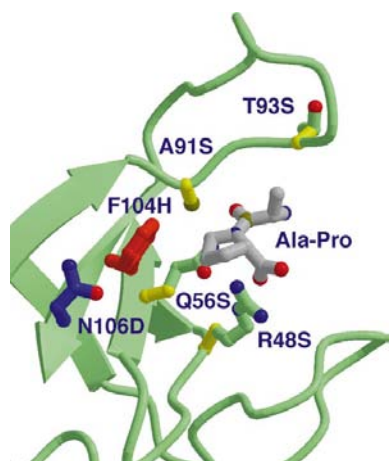
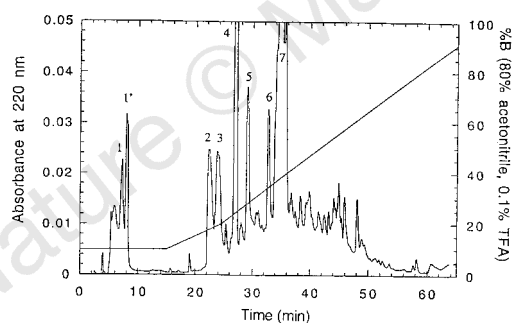


Figure 1 Enlarged view of the active site of an *E. coli* cyclophilin (PDB entry 1LOP²) illustrating the location of the Ala-Pro binding site and the enzyme residues selected for subsequent substitutions. In yellow are the C α -C β bonds of four residues (R48, Q56, A91 and T93) which were converted into serine during the first round of mutagenesis. In red and blue are Phe 104 and Asn 106 subsequently converted into His and Asp, respectively. The drawings were made using MOLSCRIPT²³ and Raster3D²⁴.



Peak number	N-terminal sequence	Measured mass	Theoretical mass	Fragment boundaries
2	LE*HNQSSQ	1,169.75 $\pm$ 0.82	1,169.559	1-10
3	PGIKLN**TTD...	1,883.54 $\pm$ 0.96	1,884.066	47-61
4	PGETN*YKKV...	3,149.08 $\pm$ 0.53	3,149.740	18-42
5	PGETN*YKKVW...	5,544.23 $\pm$ 0.51	5,544.071	18-61
6	PGETN*YKKV...	3,549.62 $\pm$ 0.56	3,549.918	18-46
7	LECHNQSSQPT...	7,678.75 $\pm$ 0.88	7,677.162	1-61

Figure 2 Degradation of a protein by cyproase I. The major products resulting from enzymatic action of ECypP on a snake curaremimetic toxin (see Methods for experimental details) were separated by HPLC (top). Eight fractions whose absorbance (A_{220}) is higher than 0.15 were numbered from 1 to 7. They were analysed by Edman sequencing and electron spray mass spectrometry with the results given in the table (bottom). Fractions 1 and 1' were non-peptidic compounds. The reaction yield was 25% as determined from the proportion of remaining non-hydrolysed toxin (peak 7). Cleavages occurred before Pro 11 (peak 2), Pro 18 (peaks 4, 5 and 6), Pro 43 (peak 4) and Pro 47 (peaks 3 and 6). An alkylated cysteine preceded Pro 18, indicating the relative permissivity of the cyproase I for large groups at the S1 site. For Pro 18 and 47, residues at the S'2 site were, respectively, Gly and Pro. Cleavage at non-proline residues cannot be excluded but must be marginal because no corresponding peptides were detected.

tion in cyclophilin-peptide complexes²⁻⁴ and a *trans* conformation in a Gly-Pro bond of a HIV1-gag hCypA complex¹². Also, although no *trans* X-Pro bonds (X being different from Gly) have ever been observed in crystal complexes, such conformers are most likely to be the main substrates of cyclophilins⁴. Therefore, the corresponding mutants were produced and their peptidase activity was measured using a chromophoric dipeptide furylacryloyl-alanylproline (fa-AP). The kinetic parameters were determined using both a spectrophotometric assay and a high-performance liquid chromatography (HPLC)-based method. The two techniques gave similar results. Table 1 shows that the wild-type ECypP, ECypP-R48S and ECypP-T93S had no detectable X-Pro peptidase activity ($k_{cat}/K_m < 10^{-3} s^{-1} M^{-1}$). In contrast, ECypP-Q56S and ECypP-A91S efficiently cleaved the substrate, the mutant A91S being more than 10^4 -fold more potent than the mutant Q56S. Because we found that spontaneous hydrolysis of an fa-Ala-Pro bond occurs at a rate ($k_{uncat} = 4.8 \pm 0.9 \times 10^{-9} s^{-1}$) that is similar to that of other peptide bonds¹³, our data demonstrate that the simple introduction of S91 in ECypP ($k_{cat} = 4.4 \pm 1.2 \times 10^{-2} s^{-1}$) enhanced by roughly 10^7 -fold the rate constant of X-Pro bond cleavage. Therefore, both location and orientation of S91 and the pre-existing environment are favourable for ECypP to express a peptidase activity.

Although there is no evidence to indicate that S91 in ECypP behaves like the nucleophile of serine proteases, we further introduced a histidine and an aspartic acid in ECypP-A91S to attempt to generate a catalytic triad-like machinery. ECypP-Q56S was not further investigated. A modelling study suggested F104 as the residue to be mutated into histidine because the imidazole group centre would be at 5.4 Å from the β -carbon of S91, as compared to 4.2 Å in a catalytic triad. ECypP-A91S-F104H was marginally more efficient than ECypP-A91S, with a 3.5-fold higher k_{cat} and a slight increase in the Michaelis constant ($K_m = 1.9$ mM).

Further modelling studies suggested that for the constraints of a catalytic triad to be approximately respected, an aspartic acid should preferably be introduced at position 106. ECypP-A91S-F104H-N106D was 20–23-fold more efficient than ECypP-A91S and ECypP-A91S-F104H, whereas the Michaelis constant remained similar to that of the double mutant. The k_{cat} of the triple mutant was nearly 100-fold higher than that of ECypP-A91S. In addition, its rate enhancement (k_{cat}/k_{uncat}) was about 10^9 , a value comparable with those observed for a number of natural enzymes¹⁴. Furthermore, the *cis-trans* isomerase activity of the wild-type ECypP vanished upon introduction of the mutations A91S, F104H and N106D (data not shown), probably as a result of a lower hydro-

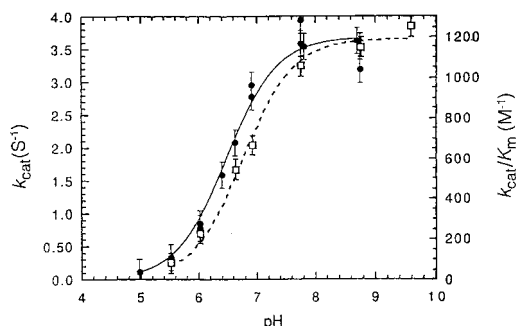


Figure 3 pH-rate profile of cyproase I. Hydrolysis of the model peptide fa-Ala-Pro was monitored spectrophotometrically in the presence of 1 and 100 nM enzyme, for determining, respectively, k_{cat} (●, solid line) and k_{cat}/K_m (□, dashed line). All measurements were performed in a mixed buffer consisting of 20 mM citrate phosphate borate at various pH levels. In this buffer, values for k_{cat} and k_{cat}/K_m measured at pH 8 were, respectively, $3.94 \pm 0.19 s^{-1}$ and $1,192 \pm 59.3 s^{-1} M^{-1}$, that is, 98.6% and 71.2% of those measured in the optimal buffer (Table 1). All values were corrected from the non-enzymatic hydrolysis which varied with pH. To calculate pK, both sets of data points were fitted to a single group titration model $y = y_{max}/(1 + 10^{pK-pH})$.

Table 1 Kinetic parameters for ECypP variants

	Efficiency k_{cat}/K_m ($s^{-1} M^{-1}$)	K_m ($\times 10^{-3} M$)	k_{cat} (s^{-1})	Rate enhancement k_{cat}/k_{uncat}	Proficiency ($k_{cat}/K_m)/k_{uncat}$ (M)
WT	$<10^{-3}$	ND			
R48S	$<10^{-3}$	ND			
Q56S	$5.8 \pm 0.8 \times 10^{-3}$	ND			
A91S	73.1 ± 2.2	0.6 ± 0.15	0.044 ± 0.012	0.9×10^7	1.52×10^{10}
T93S	$<10^{-3}$	ND			
A91S-F104H	81.8 ± 3.3	1.9 ± 0.10	0.155 ± 0.015	3.23×10^7	1.70×10^{10}
A91S-F104H-N106D	1675 ± 12	2.4 ± 0.09	4.0 ± 0.2	8.33×10^8	0.35×10^{12}

First-order kinetics were determined spectrophotometrically for the single point variants and using both the spectrophotometric and the HPLC assays for the triple mutant in the best buffer system found (20 mM HEPES, pH 8.0). K_m and k_{cat} were deduced from the Wool-Augustinsson-Hofstee plot: v versus v/S (ref. 25).

phobicity in the enzyme active site¹⁵. Therefore, the three mutations sufficed to convert the peptidyl-prolyl isomerase ECypP into an efficient X-Pro peptidase. We called this new enzyme cyproase I.

We investigated the capacity of cyproase I to cleave the polypeptide chain of a snake curaremimetic toxin which contains 61 amino acids and four disulphide bonds, and includes five proline residues. The native toxin structure is resistant to proteolytic enzymes (such as cathepsins B and D¹⁶) and, as expected, cyproase I did not hydrolyse it. After reduction of the disulphide bonds and modification of the free cysteines, the toxin unfolds¹⁷ and hence becomes cleavable by cyproase I. The peptides resulting from endoproteolysis were isolated by HPLC and characterized by Edman sequencing and electron spray mass spectrometry analysis (Fig. 2). Cleavages were observed before proline residues number 11, 18, 43 and 47. No cleavage between Pro-11 and Pro-12 was detected. Clearly, cyproase I not only cleaves the model fa-AP peptide but also displays an X-Pro-specific endoproteolytic activity. This enzyme may be of interest for future protein mapping.

How does this artificial proteolytic enzyme work? At present, we have no definite answer but a number of elements shed some light on this question. Thus, in the presence of 4-(2-aminoethyl)-benzenesulphonyl fluoride (AEBSF), cyproase I was completely inactivated within seconds (not shown), suggesting the involvement of a nucleophile in the catalytic mechanism. We suggest that S91 corresponds to this nucleophile residue, because introduction of the single mutation A91S was sufficient to generate an efficient proteolytic activity. Such a situation is not uncommon because some enzymes, including β -lactamases¹⁸ and penicillin acylase¹⁹, possess a nucleophilic serine in their active site despite having no catalytic triad. The residues that are responsible in ECypP-A91S for the nucleophilicity of S91 remain to be identified. What is perhaps more striking in our observation is that introduction of H104 and D106 further increased the efficiency of catalysis by a factor of 23. Although this value is substantially weaker than the corresponding effect observed in subtilisin¹⁰, the data suggest a synergistic action between H104, D106 and S91. That a histidine residue participates directly in the catalytic mechanism of cyproase I is supported by the observation that a group titrates with a pK equal to 6.74 ± 0.16 and 6.74 ± 0.15 , when measuring the k_{cat} and k_{cat}/K_m , respectively, versus pH (Fig. 3). Therefore the catalytic triad of cyproase I might share functional and structural similarities with conventional catalytic triads of serine proteases⁷. On the other hand, the examination of the structure of ECypP-peptide complex² suggested that the dyad H104-D106 might play the role of a general base, as in phospholipase A₂ (ref. 20). In this case, S91 would behave as an oxyanion hole motif.

There has been a constant demand for proteases with new specificity and protein engineering has been faced with this considerable challenge. Our results indicate that an alternative strategy based on modification of pre-existing enzymatic sites may lead to efficient and quite new activities. □

Method

The plasmid pJLEC-2B encoding *E. coli* periplasmic cyclophilin²¹ was kindly

provided by C. T. Walsh (Harvard Medical School, Boston). Site-directed mutagenesis was performed using Stratagene's QuickChange protocol. Cloning, sequencing and expression were made in the same plasmid construct and host strain (XL1-Blue, Stratagene). Protein production and purification of mutant proteins were achieved essentially as described²¹, with minor modifications regarding chromatography supports. An additional gel filtration on Superdex 75 (Pharmacia Biotech) was added for mutants displaying peptidase activity.

The peptidase activity of mutants was first tested by a spectrophotometric assay derived from ref. 22. The furylacryloyl-alanylproline (fa-AP) substrate was synthesized from 3-(2-furyl)acryloyl-alanyl-*N*-hydroxysuccinimide ester and *H*-proline-2-chlorotriyl chloride resin (Bachem). A molar extinction coefficient of $13,400 \text{ cm}^{-1} \text{ M}^{-1}$ at 300 nm was measured for fa-AP and its conversion was followed from the differential absorption spectra between fa-AP and fa-A at 324 nm ($\Delta\epsilon = -1,576 \text{ cm}^{-1} \text{ M}^{-1}$). Mutant cyclophilin concentrations ranged from 50 ng ml^{-1} to $5 \mu\text{g ml}^{-1}$ in the assays whereas substrate concentrations were in the range 0–1 mM. Kinetic parameters were also determined from HPLC analysis of the reaction products on a μ RPC-C2/C18 column (Pharmacia) at 0.3 ml min^{-1} , fa-Ala and fa-AP eluting, respectively, at 27.6 and 31.2% acetonitrile. Uncatalysed rate of fa-AP hydrolysis was determined by using HPLC to follow the evolution of the ratio of fa-AP/fa-A in a 20 mM fa-AP solution at 37 °C in 100 mM Tris-HCl buffer at pH 7.5 over a period of 45 days.

Toxin α from *Naja nigricollis* was reduced by 100 mM DTT, dialysed against 0.1 M acetic acid, and the eight cysteines were alkylated with 0.1 M *N*-ethylmaleimide in 50 mM Tris-Cl, pH 8.0. Hydrolysis was performed at 0.15 mM toxin in 20 mM HEPES, pH 7.0 in the presence of 0.11 μM cyproase I for 2 h at 37 °C. The products were lyophilized, resuspended in 0.1% TFA at 1 mg ml^{-1} and characterized by HPLC (C18 Bondapak 3 μ). Electron spray mass spectrometry analyses were performed at the mass spectrometry facility in DCC/SPEA, Saclay (H. Virelizier).

Received 4 June; accepted 13 October 1997.

1. Clubb, R. T., Ferguson, S. B., Walsh, C. T. & Wagner, G. Three-dimensional solution structure of *Escherichia coli* periplasmic cyclophilin. *Biochemistry* **33**, 2761–2772 (1994).
2. Konno, M., Ito, M., Hayano, T. & Takahashi, N. The substrate-binding site in *Escherichia coli* cyclophilin A preferably recognizes a *cis*-proline isomer or a highly distorted form of the *trans*-isomer. *J. Mol. Biol.* **256**, 897–908 (1996).
3. Zhao, Y. & Ke, H. Mechanistic implication of crystal structures of the cyclophilin-dipeptide complexes. *Biochemistry* **35**, 7362–7362 (1996).
4. Zhao, Y. & Ke, H. Crystal structure implies that cyclophilin predominantly catalyzes the *trans* to *cis* isomerization. *Biochemistry* **35**, 7356–7362 (1996).
5. Kirby, A. J. Enzyme mechanisms, models, and mimics. *Angew. Chem. Int. Ed. Engl.* **35**, 707–724 (1996).
6. Matthews, B. W., Craik, C. S. & Neurath, H. Can small cyclic peptides have the activity and specificity of proteolytic enzymes. *Proc. Natl Acad. Sci. USA* **91**, 4103–4105 (1994).
7. Fersht, A. *Enzyme Structure and Mechanism* (W. H. Freeman, New York, 1985).
8. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature* **337**, 476–478 (1989).
9. Galat, A. & Metcalfe, S. M. Peptidylproline *cis/trans* isomerases. *Prog. Biophys. Mol. Biol.* **63**, 69–119 (1995).
10. Carter, P. & Wells, J. A. Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564–568 (1988).
11. Bode, W. & Huber, R. Natural protein proteinase inhibitors and their interaction with proteinase. *Eur. J. Biochem.* **204**, 433–451 (1992).
12. Gamble, T. R. et al. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
13. Kahne, D. & Still, W. C. Hydrolysis of a peptide bond in neutral water. *J. Am. Chem. Soc.* **110**, 7529–7534 (1988).
14. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
15. Stein, R. L. Mechanism of enzymatic and nonenzymatic prolyl *cis-trans* isomerization. *Adv. Prot. Chem.* **44**, 1–24 (1993).
16. Maillere, B. et al. Immunogenicity of a disulphide-containing neurotoxin: presentation to T-cells requires a reduction step. *Toxicon* **33**, 475–482 (1995).

17. Ménez, A., Monteny-Garestier, T., Fromageot, P. & Hélène, C. Conformation of two homologous neurotoxins. Fluorescence and circular dichroism studies. *Biochemistry* **19**, 5202–5208 (1980).
18. Strynadka, N. C. J. *et al.* Molecular structure of the acyl-enzyme intermediate in β -lactam hydrolysis at 1.7 Å resolution. *Nature* **359**, 700–705 (1992).
19. Duggleby, H. J. *et al.* Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373**, 264–268 (1995).
20. Scott, D. L. *et al.* Interfacial catalysis: the mechanism of phospholipase A2. *Nature* **250**, 1541–1546 (1990).
21. Liu, J. L. & Walsh, C. T. Peptidyl-prolyl *cis-trans* isomerase from *Escherichia coli*: A periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. *Proc. Natl Acad. Sci. USA* **87**, 4028–4032 (1990).
22. Plummer, T. H. Jr & Kimmel, M. T. An improved spectrophotometric assay for human carboxypeptidase N1. *Anal. Biochem.* **108**, 348–353 (1980).
23. Kraulis, P. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
24. Merritt, E. A. & Murphy, M. Raster3D version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
25. Segel, I. H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems* (John Wiley, New York, 1993).

Acknowledgements. We thank J. Janin, F. Lederer and G. Robillard for useful and critical reading of the manuscript.

Correspondence and requests for materials should be addressed to E.Q. (e-mail: eric.quemeneur@cea.fr).

Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal

James R. Kiefer*, Chen Mao*, Jeffrey C. Braman† & Lorena S. Beese*

* Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, North Carolina 27710, USA

† Stratagene, 11011 N. Torrey Pines Rd, La Jolla, California 92017, USA

DNA polymerases copy DNA templates with remarkably high fidelity, checking for correct base-pair formation both at nucleotide insertion and at subsequent DNA extension steps^{1–3}. Despite extensive biochemical, genetic and structural studies^{2,4}, the mechanism by which nucleotides are correctly incorporated is not known. Here we present high-resolution crystal structures of a thermostable bacterial (*Bacillus stearothermophilus*) DNA polymerase I large fragment⁵ with DNA primer templates bound productively at the polymerase active site. The active site retains catalytic activity, allowing direct observation of the products of several rounds of nucleotide incorporation. The polymerase also retains its ability to discriminate between correct and incorrectly paired nucleotides in the crystal. Comparison of the structures of successively translocated complexes allows the structural features

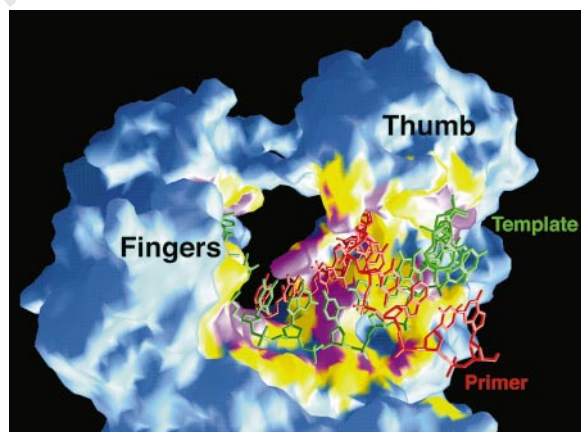


Figure 1 Structure of the *Bacillus* fragment with duplex DNA bound at the polymerase active site. The *Bacillus* fragment molecular surface is coloured according to its proximity to the DNA, with all points less than 3.5 Å coloured magenta, between 3.5 and 5.0 Å yellow, and greater than 5 Å blue. Bound water molecules were not included in this calculation.

for the sequence-independent molecular recognition of correctly formed base pairs to be deduced unambiguously. These include extensive interactions with the first four to five base pairs in the minor groove, location of the terminal base pair in a pocket of excellent steric complementarity favouring correct base-pair formation, and a conformational switch from B-form to underwound A-form DNA at the polymerase active site.

We have determined at 1.8 Å resolution the co-crystal structures of the large (relative molecular mass (M_r) 67K) carboxy-terminal fragment⁵ of the thermostable *Bacillus stearothermophilus* DNA polymerase I complexed with DNA primer templates bound at the polymerase active site. We believe that these structures provide the most detailed view of any polymerase DNA complexes yet determined (Fig. 1). The polymerase active-site region of the fragment shows extensive structural (0.65-Å root mean square deviation (r.m.s.d.) of C α atoms) and sequence homology (49% identity) to the analogous Klenow fragment⁶ from *Escherichia coli*, which has been the model system for determining polymerase structure and function for over a decade^{2,4}. Furthermore, the DNA duplex binds to the polymerase domain through an extensive network of hydrogen bonds, ion pairs, and van der Waals contacts, involving more than 40 residues which are highly conserved in polymerases of the PolI family. In particular, the residues located in the DNA polymerase active site itself, and which have been shown to be important for catalysis in the Klenow fragment^{7,8}, are invariant in the *Bacillus* fragment as well as the other members of the PolI family⁹. Preliminary characterization of the *Bacillus* fragment⁵ suggests that its enzymatic and biochemical activities are also similar to those of the Klenow fragment, although the *Bacillus* fragment is more processive, incorporating 112 nucleotides compared with 7.7 nucleotides for the Klenow fragment⁷.

Bacillus-fragment co-crystals of a DNA duplex of nine base pairs (bp) with an AGAGA-5' template overhang sequence were incubated with solutions containing the ddTTP nucleotide complementary to the first position in the overhang. The 1.9-Å resolution structure of the resulting complex clearly showed that the duplex DNA region increased from nine to ten base pairs (Fig. 2). At this resolution the DNA sequence could be read without ambiguity from the electron density, and revealed that a newly synthesized A:T (template:primer) base pair was bound at the active site in a position previously occupied by the G:C base pair of the initial substrate, indicating that the added nucleotide had been incorporated into the DNA. No density was apparent for the 3'-OH of the sugar at the new terminus, as expected because it had become a dideoxyribose. The ($F_o - F_c$) difference maps showed a pattern of peaks and holes around the DNA bases, indicating that the sequence had translocated by one base pair after incorporation. To investigate whether more than one catalytic turnover could occur in the crystal,

Table 1 Summary of DNA helical parameters

	<i>Bacillus</i> fragment 11-bp complex	B-DNA	A-DNA
Mean twist (°)	33	36	33
Mean rise (Å)	3.2	3.4	2.6
Roll (°)	3.5	0	6
Inclination (°)	1.8	-6	2.6
X-displacement (Å)	-1.4	0.2	-4.5
Sugar pucker	C2'/C3'-endo*	C2'-endo	C3'-endo
Minor-groove width/depth (Å)			
Mean	8.0/3.5	5.7/7.5†	11.0/2.8†
Distal end	7.0/4.3		
Active-site end	10.2/1.1		
Major-groove width/depth (Å)			
Mean	12.05/5.5	11.7/8.5†	2.7/13.5†
Distal end	9.6/5.8		
Active-site end	14.8/0.6		
Mean helical diameter (Å)	19.5	19	23

Bacillus fragment-DNA complex helical parameters were compiled using CURVES version 5.1 (ref. 28). Typical values for B-DNA and A-DNA are provided for comparison²⁹.

* Sugar pucker is C3'-endo for the first 4bp of DNA except the 3'-primer terminus. That residue and all others are C2'-endo.

† From ref. 30.

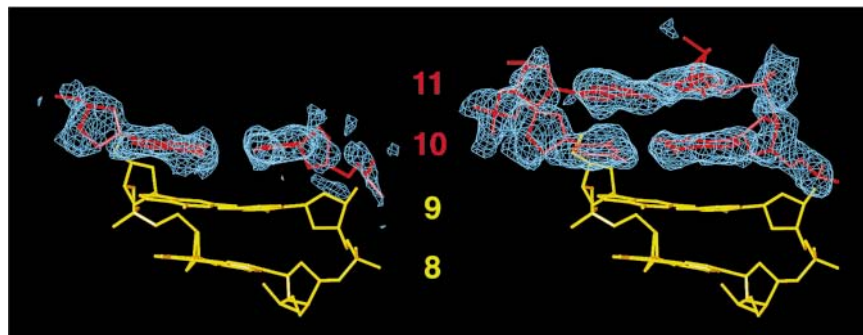


Figure 2 Electron-density maps showing catalysis in the crystal. Difference electron-density maps of DNA duplexes distal to the active site (base pairs numbered from 3' terminus of the primer strand) reveal translocation resulting from nucleotide incorporation. Addition of ddTTP to complex I (5'-AGAGA template overhang) results in single-nucleotide incorporation (left); addition of dATTP to complex II (5'-TATT overhang) incorporates two nucleotides (right). $F_o - F_c$ difference electron-density maps were calculated using refined 9-bp *Bacillus* fragment-DNA complex model phases and experimental amplitudes from nucleotide-soaked crystals.

co-crystals with the same 9-bp duplex and a different single-stranded template overhang (TTAT-5') were incubated with dATP. The 1.8-Å resolution structure revealed that this time two new T:A base pairs were added (Fig. 2), occupying the positions of the G:C and C:G base pairs in the initial substrate, which had translocated. No compelling density was apparent for a third dATP incorporation, which would have resulted in the formation of an A:A mismatch. To further investigate whether the polymerase retained the ability to discriminate against the incorporation of a mismatched base pair, TTAT-5' template overhang substrate complexes were incubated with each of the other three nucleotide triphosphates. No significant incorporation was observed, even when the soak times were extended. These experiments therefore unambiguously demonstrate that the polymerase retains catalytic activity in the crystals, successively incorporating nucleotides into the primer DNA strand, translocating the DNA product, and discriminating between correct and incorrect base-pair formation.

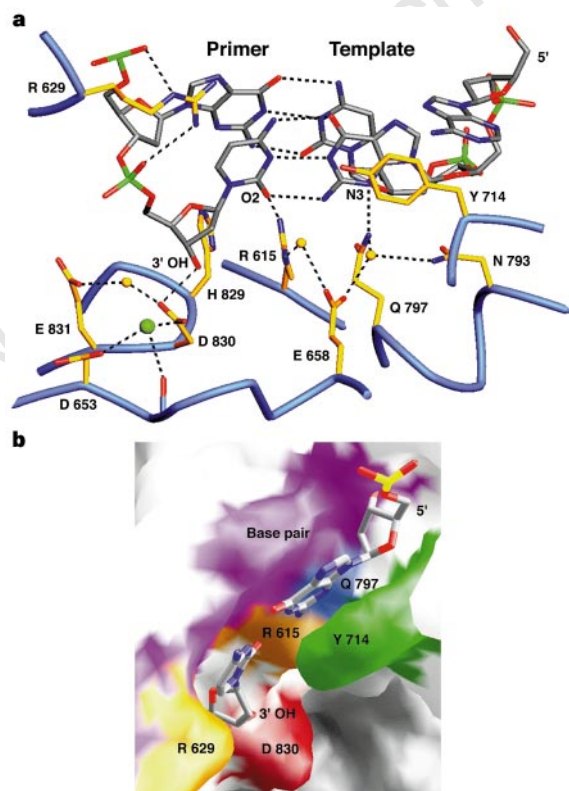


Figure 3 Polymerase active site. **a**, All residues shown are strictly conserved in enzymes of the Pol I family. Yellow spheres represent water molecules, the green sphere a Mg^{2+} ion. Klenow fragment homologues of residues shown are listed in brackets: 615(668), 629(682), 653(705), 658(710), 714(766), 793(854), 797(849), 829(881), 830(882) and 831(883). **b**, Molecular surface representation of the pocket surrounding the first base pair at the active site. The primer is on the left.

Polymerases must provide sequence-independent interactions with their DNA substrate, yet retain the specificity to distinguish correctly paired bases from mismatches. We can observe the detailed interactions between the protein and three different DNA sequences formed by one or two rounds of successive nucleotide incorporation. One of the remarkable features of these complexes is the extensive sequence-independent interactions in the minor groove of the first four base pairs extending from the 3' primer terminus, which form hydrogen bonds to the N3 positions of purine bases and O2 position of pyrimidine bases through highly conserved protein side chains or oriented water molecules anchored to protein side chains. Only these two groups present the same, two-fold symmetric pattern of hydrogen-bond acceptors in both G:C and A:T base pairs¹⁰. The major groove, with its sequence-specific pattern of hydrogen-bond donors and acceptors, which form the primary means of recognition for many sequence-specific DNA-binding proteins^{11,12}, does not contact the protein and is solvent accessible. The interactions in this minor-groove recognition (MGR) region are independent of DNA sequence, and also provide additional specificity for correctly paired bases. This is particularly pronounced at the terminal, primer base pair in the active site itself. This base pair binds in a pocket formed by stacking interactions between Tyr 714 (766) (the corresponding residue in the Klenow fragment is given in brackets) and the template base, by van der Waals interactions with residues of the palm subdomain that shape the floor and walls of the pocket, and by the second base pair (Fig. 3). Thus, in addition to hydrogen bonds to the strictly conserved residues Arg 615 (668) and Gln 797 (849), this pocket provides greater specificity for correctly formed paired nucleotides, by steric complementarity not observed at the other three positions of the MGR region. A mismatch could not bind with the same geometry as a correctly formed base pair¹³, resulting in misalignment of the 3'-OH primer terminus, thereby preventing incorporation of the next nucleotide. In the Klenow fragment, several of the residues homologous to those that interact with the first base pair have been found by site-directed mutagenesis to have a dramatic effect on fidelity and k_{cat} (refs 7, 8, 14, 15). Most striking are mutations that remove the terminal stacking interaction with the template base (Y766A or Y766S of the Klenow fragment), which results in a significant decrease in the fidelity of replication^{14,15}, whereas the neutral change Y766F does not¹⁴.

The extensive minor-groove interactions are made possible by significant underwinding of the DNA (Table 1). The DNA outside the MGR region almost adopts a regular B-form conformation with its characteristic C2'-endo sugar pucker, clearly observable at 1.8-Å resolution. Pronounced conformational transitions occur at the star of the MGR region. The sugar pucker switches to the C3'-endo conformation characteristic of A-form DNA. This is accompanied by a monotonic widening of the minor groove, ranging from 7.0 to 10.2 Å at the active site (Table 1). Even more striking is that the minor groove becomes extremely shallow (1.1 Å at the active site), allowing easy access of the protein side chains to the edges of the bases. Beyond the MGR region, in the B-form region, there are no

Table 2 Summary of crystallographic and refinement data

X-ray diffraction statistics	9-bp DNA	Data set 9-bp DNA + ddTTP	9-bp DNA + ddTTP
Resolution limit (Å)	1.8	1.9	1.8
No. reflections, unique/total	73,375/253,450	63,997/326,641	72,444/314,981
Mean I/σ , overall/outer shell	17.5/5.1	26.4/6.1	18.9/5.6
Completeness (%), overall/outer shell	91.7/60.1	92.8/75.4	90.6/76.6
R_{sym} (%), overall/outer shell	4.4/16.3	6.1/18.7	4.7/13.3
Refinement ($F \geq 2\sigma_F$)			
Resolution range (Å)	6.0–1.8	6.0–1.9	6.0–1.8
No. unique reflections	69,994	61,255	68,602
R -factor (%)	22.1	21.3	20.8
R_{free} (%)	26.6	27.4	26.7
No. solvent molecules	527	497	606
R.m.s. bond lengths (Å)	0.010	0.012	0.010
R.m.s. bond angles (°)	1.478	1.453	1.437
R.m.s. improper angles (°)	1.054	1.026	1.010
R.m.s. dihedral angles (°)	24.82	25.33	25.09

$R_{\text{sym}} = (\sum |I - \langle I \rangle|) / (\sum I)$, where $\langle I \rangle$ is the average intensity of multiple measurements. R -factor and $R_{\text{free}} = (\sum |F_{\text{obs}} - F_{\text{calc}}|) / (\sum |F_{\text{obs}}|)$. R_{free} was calculated over 5% of the amplitudes not used in refinement. R.m.s. deviations reported include both protein and DNA residues.

interactions between the protein and the nucleotide bases: all protein–DNA interactions involve the sugar–phosphate backbone, and are therefore also independent of sequence. The point of the MGR boundary is preserved in all the complexes: as a base pair translocates from the fourth to the fifth position in the successively formed complexes its sugar pucker undergoes a conformational transition. Mismatches in the first three to four base pairs from the primer terminus cause the homologous T7 DNA polymerase to stall¹, suggesting that the MGR region is tightly coupled to DNA translocation.

The location of the 3'-OH terminus is well positioned in the active site for catalysis (Fig. 3a). It forms a hydrogen bond to Asp 830 (882), an invariant residue in all polymerases¹⁶, and is essential for activity in the Klenow fragment as well as the *Bacillus* fragment (J.C.B., unpublished observation). In forming the hydrogen bond between the 3'-OH and Asp 830, the sugar of the primer nucleotide is observed to adopt a C2'-endo conformation, making it the one exception to the observed A-form DNA conformation of the first four bases. For the 3'-OH of the sugar to be in line to attack the α -phosphate of an incoming nucleotide triphosphate, it is

necessary only to switch the sugar pucker to the A-form C3'-endo. Although we have not yet captured such an incoming dNTP in our complexes, the location of a modelled dTTP residue (Fig. 4) is in good agreement with biochemical and mutagenesis studies the Klenow fragment^{2,7,8}. A divalent metal ion observed in our crystals to be coordinated by invariant residues Asp 653 (705) and Asp 830 stabilizes the α -, β - and γ -phosphates of the modelled dNTP. The modelled α -phosphate of the dTTP is 1.7 Å from the 3'-OH, and both could interact with a second metal ion (not observed in our crystals). This position of the dTTP is consistent with both the two-metal-ion mechanism for catalysis^{2,17} and direct activation of the 3'-OH by Asp 830 (refs 8, 18). The 3'-OH of the dTTP sugar is positioned over Glu 658 (710) and adjacent to Phe 710 (762), which affect sugar specificity when mutated^{19,20}.

A surprising feature of the active site is that Tyr 714 is in exactly the position predicted for the first nucleotide of the 5' template-strand overhang (Figs 3, 4). As a consequence, the template-strand overhang takes a sharp turn ($\sim 90^\circ$) relative to the axis of the DNA duplex where it binds in a shallow pocket formed between the end of the O helix and the beginning of the O₁ helix (residues 712–724). Consequently, it is impossible for an incoming dNTP to form a normal Watson–Crick base pair with the template in this position. A conformational change involving movement of the C terminus of the O helix in addition to the tyrosine is therefore required for the template base to become accessible for Watson–Crick base-pair formation, given that catalysis occurs with the primer in its observed position. Such a change may correspond to the slow kinetic step observed to separate dNTP binding and covalent incorporation in the homologous T7 and Klenow-fragment DNA polymerases^{21,22}.

These complexes represent a snapshot of the multistep polymerase reaction and, because the polymerase retains catalytic activity in the crystals, allow us to show unambiguously that the observed complexes are part of the reaction pathway. The observed resting state of the reaction in the crystal probably corresponds to the product complex. Other steps in the pathway may now also be established using dNTP analogues and Laue crystallography. □

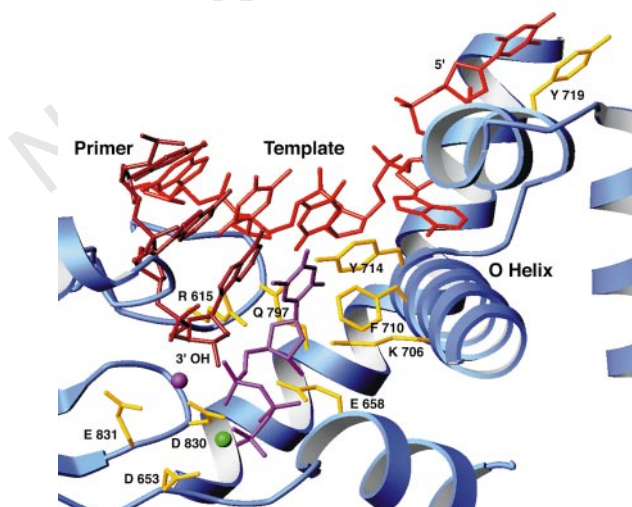


Figure 4 Polymerase active site with observed DNA and modelled dTTP. The position of dTTP (violet) was based on the β -polymerase complex¹⁸, adjusted such that the base ring stacks with the primer and one oxygen from each phosphate group was within 3 Å of the observed metal ion (gold). The sugar pucker of the primer terminus was made C3'-endo, which shifted its 3'-OH to within 1.7 Å of the modelled α -phosphate of the dTTP. A second metal ion (violet) was modelled to be within 3 Å of the 3'-OH of the primer, the α -phosphate group, and residues Asp 830 and Glu 831. The observed 5' template overhang cannot accept an incoming dNTP without a conformational change of the O helix.

Methods

Co-crystallization of the *Bacillus* fragment with DNA primer template substrates. The protein of the *Bacillus* fragment was purified from an *E. coli* overexpression system⁵. In separate experiments, two sets of complementary oligonucleotides (complex I, 5'-GCATGATGC-3' and 5'-AGAGAGCATCATGCp-3'; complex II, 5'-TATTGCATGATGC-3' and 5'-TATTGCATCATGC-3') were annealed in 10 mM cacodylate buffer (pH 7.0), 50 mM NaCl, 10 mM MgSO₄, 0.5 mM EDTA. The resulting DNA substrates consist of an identical nine-duplex base-pair region and a 5' single-stranded overhang (complex I, 5'-AGAGA; complex II, 5'-TATT) forming the template strand for the polymerase. The optimal duplex length was deduced from a different 13-bp

complex of *Bacillus* fragment and DNA with a three-nucleotide 5' template overhang in which the DNA bound in symmetric orientations with both blunt and recessed 3' ends at the polymerase active site (C.M. and L.S.B., unpublished observation). Based on this observation, in complex I the 3' end of the template strand was phosphorylated to fix the orientation of the DNA. In complex II, the pseudo-symmetric sequence presents an identical template overhang in both orientations. DNA was incubated with a 10 mg ml⁻¹ protein solution in a 3:1 molar ratio. Crystals of the complex of *Bacillus* fragment and DNA were formed by hanging-drop vapour diffusion⁵ and transferred into a cryoprotectant solution containing 24% (w/v) sucrose before data collection.

Data collection and structure determination. Diffraction data were collected at 98 K using a RAXIS-IIC detector (Molecular Structure) on a Rigaku rotating anode X-ray generator. The crystals of the complex belong to the orthorhombic space group of $P2_12_12_1$ with unit cell dimensions $a = 86.2 \text{ \AA}$, $b = 93.2 \text{ \AA}$ and $c = 106.4 \text{ \AA}$, and contain one molecule per asymmetric unit. Data were processed using DENZO and SCALEPACK²³ (Table 2). Rigid-body refinement²⁴ of the apoenzyme coordinates⁵ provided the initial phases for the 9-bp *Bacillus* fragment–DNA complex data. Inspection of ($F_o - F_c$) and ($2F_o - F_c$) difference electron-density maps showed striking helical density for 9 bp of duplex DNA. The nucleotides were fit into this density using O²⁵ and the structure was refined with X-PLOR²⁴ using the new DNA force field and topology parameters²⁶ (Table 2).

The thumb subdomain (residues 507–585) is in a significantly different conformation from the apoenzyme structure and was rebuilt into difference electron-density maps updated during refinement. Residues 548–553, which were disordered in the apoenzyme structure, are seen in the *Bacillus* fragment–DNA complex and interact with the phosphate backbone (Fig. 1). The entire *Bacillus* fragment protein (residues 297–876) is fit in the electron density, except for 12 residues at the N terminus which are disordered. The average protein B-factor is 21 Å², and 26 Å² for the DNA. Density for the first two nucleotides of the template overhang at the active site is apparent, although the remaining 5' single-stranded nucleotides are disordered.

Catalysis in the crystal. Complex I or II co-crystals were transferred into a stabilizing solution (2.5% MPD, 100 mM MES buffer, pH 5.8, 60% (v/v) saturated ammonium sulphate) containing 33 mM ddTTP and 50 mM MnSO₄ (complex I) or 22.5 mM dATP and 50 mM Mg SO₄ (complex II) and incubated for 2–3 days at room temperature. The refined 9-bp *Bacillus* fragment–DNA coordinates were used as the initial model to phase the data from nucleotide triphosphate-soaked crystals. New electron density at the end of the DNA distal from the active site was observed in ($F_o - F_c$) and [F_o (nucleotide soaked) – F_o (unsoaked)] difference electron-density maps phased by the coordinates of the 9-bp complex (calculated using CCP4 (ref. 27)). This new density (Fig. 2) was consistent with the formation of one (ddTTP experiment) or two (dATP experiment) additional DNA base pairs. Positive and negative peaks in the ($F_o - F_c$) difference maps were also observed on the edges of the DNA base rings, indicating that the register of the initial modelled DNA sequence relative to the protein had changed. The DNA was rebuilt into the density, translocated by one base pair for the ddTTP experiment and two base pairs for the dATP experiment to fill the newly observed density at the distal end of the DNA. Additional base pair(s) representing the incorporated nucleotide(s) were modelled at the active site. Refinement of the model with the rebuilt, translocated DNA sequence caused an immediate 4% drop in the R-factor to 26% and eliminated the ($F_o - F_c$) peaks on the edges of the DNA base rings.

Received 9 September; accepted 18 November 1997.

- Johnson, K. A. Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685–713 (1993).
- Joyce, C. M. & Steitz, T. A. Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63**, 777–822 (1994).
- Echols, H. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60**, 477–511 (1991).
- Joyce, C. M. & Steitz, T. A. Polymerase structures and function: variations on a theme? *J. Bacteriol.* **177**, 6321–6329 (1995).
- Kiefer, J. R. *et al.* Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95–108 (1997).
- Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762–766 (1985).
- Polesky, A. H., Steitz, T. A., Grindley, N. D. F. & Joyce, C. M. Identification of residues critical for the polymerase activity of the Klenow Fragment of DNA polymerase I of *Escherichia coli*. *J. Biol. Chem.* **265**, 14579–14591 (1990).
- Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. F. & Joyce, C. M. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I of *Escherichia coli*. *J. Biol. Chem.* **267**, 8417–8428 (1992).

- Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787–802 (1993).
- Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804–808 (1976).
- Steitz, T. A. Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* **23**, 205–280 (1990).
- Harrison, S. C. A structural taxonomy of DNA-binding proteins. *Nature* **353**, 715–719 (1991).
- Steitz, T. A., Beese, L. S., Freemont, P. S., Friedman, J. M. & Sanderson, M. R. Structural studies of Klenow fragment: an enzyme with two active sites. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 465–471 (1987).
- Carroll, S. S., Cowart, M. & Benkovic, S. J. A mutant of DNA polymerase I (Klenow Fragment) with reduced fidelity. *Biochemistry* **30**, 804–813 (1991).
- Bell, J. B., Eckert, K. A., Joyce, C. M. & Kunkel, T. A. Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272**, 7345–7351 (1997).
- Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461–467 (1990).
- Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).
- Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. Structures of ternary complexes of rat DNA polymerase β , a DNA template–primer, and ddCTP. *Science* **264**, 1891–1903 (1994).
- Tabor, S. & Richardson, C. C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl Acad. Sci. USA* **92**, 6339–6343 (1995).
- Joyce, C. M. Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl Acad. Sci. USA* **94**, 1619–1622 (1997).
- Wong, L., Patel, S. S. & Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: Direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526–537 (1991).
- Kuchta, R. D., Benkovic, P. & Benkovic, S. J. Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry* **27**, 6716–6725 (1988).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276A**, 307–326 (1997).
- Brünger, A. T. *X-PLOR version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
- Jones, T. A., Zou, J.-Y. & Cowan, S. W. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. New parameters for the refinement of nucleic acid containing structures. *Acta Crystallogr. D* **52**, 57–64 (1996).
- Collaborative Computational Project No. 4. *Acta Crystallogr. D* **50**, 607 (1994).
- Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: Conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655–667 (1989).
- Dickerson, D. E. in *Oxford Handbook of Nucleic Acid Structure* (ed Neidle, S.) (Oxford University Press, Oxford, UK, 1997).
- Saenger, W. *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).

Acknowledgements. We thank S. Johnson and A. Chapin Rodriguez for assistance in data collection and crystallization, and H. W. Hellinga for discussions. This work was supported by grants to L.S.B. from the American Cancer Society, North Carolina Biotechnology Center, and the Searle Scholars Program.

Correspondence and requests for materials should be addressed to L.S.B. (e-mail: lsb@bcm.biochem.duke.edu). Coordinates for the *Bacillus* polymerase–DNA complexes have been deposited in the Brookhaven protein database under accession nos 2BDP, 3BDP and 4BDP (the apo *Bacillus* polymerase is 1BDP).

erratum

Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures

Pamela J. Skinner, Beena T. Koshy, Christopher J. Cummings, Ivan A. Klement, Kara Helin, Antonio Servadio, Huda Y. Zoghbi & Harry T. Orr

Nature **389**, 971–974 (1997)

Figure 4c, d was reproduced with extraneous colour obscuring the gel patterns. The correct version is shown here. □

